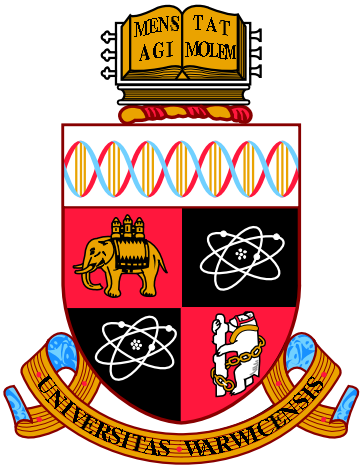


Accelerated Estimation and Inference for Heritability of fMRI Data

by

Xu Chen

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

September 2014

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

Foremost, I would like to express my deepest gratitude to my supervisor Prof. Thomas Nichols for his guidance and support throughout my PhD study.

I am also very grateful to those at the Department of Statistics and the University of Warwick for always providing help when needed.

I would also like to thank the Department of Statistics and the University of Warwick for the financial support I have received.

To my parents and my grandmother

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented in this thesis was carried out by myself, except where stated. Parts of this thesis (Chapters 3 and 4) will be submitted for publication soon.

Abstract

In this thesis, we develop some novel methods for univariate and multivariate analyses of additive genetic factors including heritability and genetic correlation.

For the univariate heritability analysis, we present 3 newly proposed estimation methods—Frequentist ReML, LR-SD and LR-SD ReML. The comparison of these novel and those currently available approaches demonstrates the non-iterative LR-SD method is extremely fast and free of any convergence issues. The properties of this LR-SD method motivate the use of the non-parametric permutation and bootstrapping inference approaches. The permutation framework also allows the utilization of spatial statistics, which we find increases the statistical sensitivity of the test.

For the bivariate genetic analysis, we generalize the univariate LR-SD method to the bivariate case, where the integration of univariate and bivariate LR-SD provides a new estimation method for genetic correlation. Although simulation studies show that our measure of genetic correlation is not ideal, we propose a closely related test statistic based on the ERV, which we show to be a valid hypothesis test for zero genetic correlation. The rapid implementation of this ERV estimator makes it feasible to use with permutation as well.

Finally, we consider a method for high-dimensional multivariate genetic analysis based on pair-wise correlations of different subject pairs. While traditional genetic analysis models the correlation over subjects to produce an estimate of heritability, this approach estimates correlation over a (high-dimensional) phenotype for pairs of subjects, and then estimates heritability based on the difference in MZ-pair and DZ-pair correlations. A significant two-sample t-test comparing MZ and DZ correlations implies the existence of heritable elements. The resulting summary measure

of aggregate heritability, defined as twice the difference of MZ and DZ mean correlations, can be treated as a quick screening estimate of whole-phenotype heritability that is closely related to the average of traditional heritability.

Chapter 1

Introduction

The objective of this thesis is to develop methods for the analysis of genetic heritability suitable for brain imaging data. Heritability measures the extent to which the inter-individual variation in twin data can be accounted for by the genetic influences; this can be considered for a single observable trait (i.e., trait heritability), as well as for multiple traits, where heritability concerns the overlapping genetic effects shared between measurable characteristics (i.e., genetic correlation between traits). Based on the data from a single trait, we construct novel heritability estimation and inference approaches with the goal of increasing computational efficiency and statistical sensitivity while controlling the type I error rate over the whole brain. We extend these results to the multivariate setting, where the genetic correlation of the paired traits can be estimated with our newly proposed method and the significance of both heritability and genetic correlation can be tested simultaneously with a test statistic that performs extremely quick and satisfactorily. Finally, during the high-dimensional multivariate heritability inference, we analyze a rapid aggregate measurement of the heritability over the entire brain volume, yielding a fast screening method useful when searching over a large collection of high-dimensional imaging phenotypes.

1.1 Background

As the most complicated structure within the central nervous system, the human brain is an organ that is highly organized to control the body's vital functions by sending, receiving, analyzing and storing information from the internal and external environments. In the human brain, there are many nerve cells including around 100 billion neurons and about 10–50 times more glial cells, where neurons receive

and transmit electrochemical signals across the brain and nervous system, and glial cells provide physical and nutritional support for neurons (Chudle, 2007). The earlier studies have revealed that the adult human brain receives 15–20% of the body's blood supply and consumes roughly 20% of the amount of whole-body energy budget for supporting brain functions, even though the brain only occupies 2% of the total body weight (Raichle, 2006).

Gaining an explicit understanding how the brain works is one of the greatest challenges scientists face. The recent developments in imaging techniques provide neuroscientists new insights into the investigation of neurophysiological and psychological diseases and allow researchers to take a closer look at the brain structures and functions. In particular, the non-invasive scanning techniques, such as magnetic resonance imaging (MRI) and positron emission tomography (PET), are imaging approaches that aid in the diagnosis of diseases and the detection of intrinsic physiological activities.

For over a decade there has been a rapidly growing interest in joint studies of imaging phenotypes and molecular genetics (Glahn et al., 2007). Imaging genetics is a multi-disciplinary research field studying the genetic influences on brain structures and functional activities using the data obtained from neuroimaging and genetics. The imaging techniques are integrated with the genetic analytical approaches, with the potential of providing greater statistical power and more effective prediction in neuroimaging (Tairyan and Illes, 2009). The recent advances in both molecular genetics and non-invasive neuroimaging provide numerous promising approaches for researchers to investigating the underlying mysteries of the human brain. Imaging genetics also allows the researchers to explore and identify the gene-related diseases before the behavioral symptoms are present, fostering the potential of patients to seek appropriate medical treatments earlier (Hariri et al., 2006).

A trait in biology can be any single feature or quantifiable measurement of an organism, and a phenotype in genetics is an individual's observable physical trait or characteristic that results from the interaction of genetic inheritance and environmental conditions. Heritability of a phenotypic trait is conceptualized as a fraction of the phenotypic variation that is attributable to the genetic influence, and summarizes how strong the genetic effects are for that phenotype (see Section 2.4 for a precise definition of heritability). The genetic effects of a trait can be attributed to a particular gene only or the aggregate of multiple genes. For the neuroimaging

traits of interest, neuroscientists want to evaluate the unknown genetic causes, and precisely measure the degree of the genetic factors, i.e., heritability. Typically, some traditional (non-imaging) phenotypes have been studied in heritability analysis of psychiatric disorders, which are mostly moderately to highly heritable, like heritability estimates of 0.83 for schizophrenia (Cannon et al., 1998) and 0.85 for bipolar affective disorder using narrow concordance (McGuffin et al., 2003). Similarly in the brain imaging studies, prior research has observed considerable genetic influence on neuroanatomy and functional neuroimaging, such as the heritability estimates of 0.88 for total gray matter volume and 0.85 for total white matter volume (Glahn et al., 2007), and the highest heritability estimates with a range of 0.40–0.65 for working memory task-related brain response (Blokland et al., 2011).

If the genetic data is available, heritability for a particular gene or a genetic marker can be measured. Without the genetic data, heritability can only be estimated by studying the individuals with varying degrees of genetic relatedness. The classic twin study is a valuable tool designed to measure the genetic effects distinct from the environmental effects common to each twin pair. Family studies, with more variation in relatedness, can also be used to explore the relative importance of genetic and environmental factors to the observable traits.

1.2 Existing Approaches

Currently there exist two commonly used approaches to estimating heritability, one based on method of moments, and another based on maximum likelihood. The so-called "Falconer's formula" (Falconer and Mackay, 1996) is a simple point estimate for heritability using twin data based on moment matching. The best-practice and commonly used likelihood-based method constructs a variance component model using a variance decomposition, which parameterizes the varying degree of covariance expected with varying relatedness between subjects; the variance parameters are estimated by maximizing the likelihood or restricted likelihood function with respect to those parameters. Hypothesis tests are generally made with likelihood ratio test, and the asymptotic standard error of the maximum likelihood estimator can be computed with the Fisher information matrix. These methods for heritability inference will be discussed in Section 3.2.

While the first neuroimaging studies measuring heritability used Falconer's method, e.g., Wright et al. (2002), likelihood-based approaches are becoming more prevalent,

with the use of variance components or structural equation modeling methods. But such methods cannot exploit the spatial nature of the imaging data to improve the statistical sensitivity, nor provide inferences corrected for searching the brain for heritability; in particular, neuroimaging researchers wish to control the rate of family-wise errors (FWE), the chance of making one or more false positives among all hypotheses in the whole brain. Although a simple Bonferroni correction is a p-value adjustment approach offering the control of FWE, it is typically quite conservative for smooth images, and thus is often not useful with brain imaging data. When feasible, permutation inference offers an exact control of false positive risk and allows the use of specialized spatial statistics while implicitly accounting for the spatial dependence. However, the maximum-likelihood-based approaches mentioned above are mostly iterative and can be too time-consuming and unreliable (i.e., can fail to converge) to allow permutations. Another, quite practical concern, is that current tools are not designed for brain imaging data, cannot compute basic spatial statistics, and can neither read nor write brain image file formats.

1.3 Thesis Organization and Main Contributions

The organization of this thesis is as follows. In Chapter 2, we review the basic principles and concepts of functional magnetic resonance imaging (fMRI). The fMRI data pre-processing and modeling approaches are also introduced in this chapter, which is followed by the description of principles, terminology and approaches related to the heritability inference, where we also present the basics of permutation and bootstrapping approaches in detail. In the subsequent chapters including Chapter 3 and Chapter 4, we propose a linear regression-based method that is new to the neuroimaging community. This method allows fast voxel-wise heritability estimation with an approximate but remarkably accurate performance. In Chapter 3, we focus on univariate heritability inference, which is developed for both voxel- and cluster-wise inferences, with its corresponding application to two real fMRI BOLD datasets. In Chapter 4, the bivariate modeling approach is proposed and evaluated with simulations, and then is further evaluated with real data. An aggregate heritability inference approach for high-dimensional multivariate heritability analysis is also investigated with the use of correlation mean difference approach in this chapter. In Chapter 5, we summarize our work in Chapter 3 and Chapter 4 and discuss our future research direction.

Chapter 2

Background on Magnetic Resonance Imaging and Heritability Analysis

Before statistically analyzing the acquired datasets, we focus on the introduction of the background information related to data acquisition, pre-processing and modeling firstly in this chapter, and then we introduce the related terminology and describe the inference approaches used in heritability analysis.

2.1 Functional Magnetic Resonance Imaging

The functional magnetic resonance imaging (fMRI) procedure, as a burgeoning neuroimaging method, uses the magnetic resonance imaging (MRI) technology to detect and measure the brain activity by means of associated changes in blood flow, volume and oxygenation. Since the 1990s, the fMRI technique has gradually become the most widely applied method for mapping brain function, owing to its non-invasiveness, excellent spatial and good temporal resolutions, and wide availability. In this section, we will briefly introduce the MRI technique, and then describe the important Blood Oxygenation Level Dependent (BOLD) response.

2.1.1 A Brief Introduction to MRI

The rapid growth of neuroimaging is supported by various imaging techniques, such as MRI, positron emission tomography (PET), electroencephalography (EEG) and magnetoencephalography (MEG), for the purpose of visualizing the brain structures and functions. Compared with EEG and MEG, which provide brain images with

high temporal resolution but relatively poor spatial localization, MRI has excellent spatial resolution to allow better visualization and reliable localization of the brain regions; in particular, it offers a means to capture both brain structure as well as function. While PET involves the injection of radioactive tracers, MRI is completely non-invasive and involves no ionizing radiation. MRI is utilized for the studies of human brain structure in vivo and has become clinically prevalent. With the use of the principles of nuclear magnetic resonance (NMR), MRI can produce high-quality cross sectional images. The NMR phenomenon is based on the quantum-mechanical properties of nuclear spin in an external magnetic field, which underlies the MRI technology.

In 1946, Bloch and Purcell discovered the phenomenon of NMR and introduced this technique independently. In 1952, Gabillard proposed the method to investigate NMR under a gradient, or spatially-varying, magnetic field for spatial discrimination of the NMR phenomenon. In 1971, Damadian's studies of rats showed that the NMR relaxation time in healthy and tumor tissues differed, providing the basis for tissue discrimination in the analysis of medical images and disease identification. Nowadays, the MRI technique has been developed and widely used as a common analytical imaging tool for chemical and physical molecular analysis.

According to the principles of quantum mechanics, if there is an odd number of protons or neutrons in the nucleus of a specified atom, this atomic nucleus has a magnetic moment and is NMR-active. A hydrogen nucleus has a single proton and zero neutrons, and therefore is NMR active. Since the hydrogen nuclei are the most abundant in the human body, making up about 60% of the human body (e.g., each water molecule consists of two hydrogen atoms and one atom of oxygen), conventional MRI obtains the magnetic signal from hydrogen nuclei (protons).

Each hydrogen proton has the magnetic characteristic of nuclear spin, and usually the axes of these spins are randomly oriented. In the presence of an external, strong and static main magnetic field (with the strength of B_0) for the hydrogen atoms, the hydrogen protons align themselves with this field. They align both parallel and anti-parallel to the applied magnetic field, producing a net magnetization M_0 obtained along with the direction of the external field. Roughly, each hydrogen proton can be thought of like a little magnet. With the influence of the external static magnetic field, these little magnets precess around the axis of the external field. The frequency of precession occurs at the resonant frequency determined by

the Larmor equation:

$$f = \gamma B_0, \tag{2.1}$$

where $\gamma = 42.58$ MHz/T is the gyromagnetic ratio of hydrogen atoms, and B_0 is as stated to be the strength of this external magnetic field in Tesla (T).

The "little magnets" of the hydrogen protons cannot be directly detected when aligned with the massive external field. To measure these protons the phenomenon of magnetic resonance is used. An oscillating magnetic field (with the strength of B_1), perpendicular to the main static field, is applied for a very short interval at the resonance (Larmor) frequency. This B_1 magnetic field is also referred to as a radio frequency (RF) field, and turning on this RF field for a brief period is known as RF pulse. After applying an RF pulse, the spinning hydrogen protons absorb energy, become excited and precess in phase and at an angle tipped away from the main B_0 field towards the B_1 field. Following the excitation, these excited hydrogen atoms will return to their equilibrium state by emitting energy at the same radio frequency and becoming relaxed gradually when the RF pulse is switched off. The process of energy exchange for the hydrogen protons by absorbing and releasing energy at the Larmor frequency is called magnetic resonance and gives rise to the MRI signal.

The magnetization M_0 can be decomposed into two orthogonal components: a longitudinal or z component M_z arisen from the main B_0 field, and a transverse component M_{xy} lying on the transverse xy plane. After the RF excitation, the longitudinal and transverse components of the magnetization return to their equilibrium values at an exponential rate, where the relaxation process can also be separated into two parts: spin-lattice relaxation and spin-spin relaxation for the longitudinal and transverse components respectively. The recovery of the longitudinal component for spin-lattice relaxation is characterized by a relaxation time constant T_1, whereas T_2, the spin-spin relaxation time constant, describes the decay process on the transverse plane, where T_1 is typically 5-10 times larger than T_2. Specifically, these two independent exponential processes of the longitudinal recovery and the transverse decay can be characterized by the temporal changes of M_z along z axis and M_{xy} within the xy plane over time t:

$$M_z(t) = M_0\Big[1 - \exp\Big(-\frac{t}{T_1}\Big)\Big],$$

$$M_{xy}(t) = M_0 \exp\Big(-\frac{t}{T_2}\Big),$$

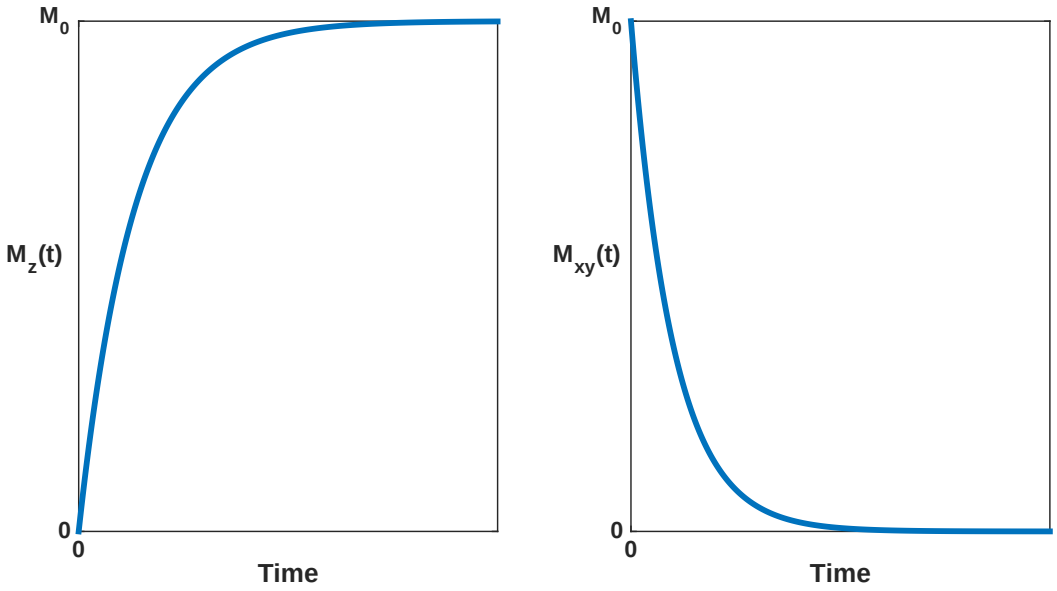


Figure 2.1: The exponential processes of longitudinal recovery of the $M_z(t)$ component (left) and transverse decay of the $M_{xy}(t)$ component (right) over time once the RF pulse is switched off.

which are depicted in Figure 2.1 (Hashemi et al., 2010). The spatial inhomogeneity of the applied static magnetic field and the interactions between the magnetized protons within the human body cause the phase dispersal, leading to a more rapid decay of the transverse component and a faster reduction of the acquired MRI signal (Hashemi et al., 2010). Hence the observed spin-spin relaxation time, denoted as T_2^*, is always shorter than T_2. The clinical success of MRI is derived from the fact that different tissues (e.g., white matter, grey matter and cerebral spinal fluid) have distinct combination of T_1 and T_2 time constants, and thus distinct image contrast.

The use of the spatially varying gradient magnetic field for spatial encoding is the fundamental of spatial localization and creation of 3D volumes with the MRI technique. The details of "spatial encoding", the process of how spatial information is induced in resonance properties of the object and subsequently reconstructed, are involved; see Weishaupt et al. (2006); Hashemi et al. (2010) for a detailed description. In brief, linearly-varying gradient fields are controlled and applied along the direction of x, y and z axes separately, which results in distinct resonance frequencies for different locations in the 3D space. The Fourier transform can be used to decode the received MRI signal so as to discriminate and identify the spatial pos-

itions associated with different resonance frequencies based on Equation (2.1). In the longitudinal direction, an 1D RF excitation pulse is applied for the separation and selection of different brain slices. For each longitudinal slice, the employment of frequency encoding and phase encoding for the adjunction of two gradient fields on the x and y axes respectively makes the localization of spatial positions on the 2D transverse xy plane feasible by converting the MRI data into the corresponding spectrum. The phase encoding on the y axis requires repetitive excitations of the selected slice for each phase difference, which is often time-consuming, and thus a fast MRI technique of echo planar imaging (EPI), proposed in 1977 by Mansfield (Mansfield, 1977), is commonly used, involving a single RF excitation pulse to obtain all information of spatial encoding.

2.1.2 Hemodynamic Response

The fMRI technique, as a specialized form of MRI, was developed to indirectly measure and image the neural activity caused by task performance or stimulus with the accompanying hemodynamic response. A number of studies have demonstrated there exists a linear relationship between neural activity and hemodynamic response (Logothetis, 2003). The procedure of fMRI is based on the different magnetic susceptibility of oxyhemoglobin and deoxyhemoglobin, and the changing concentration of deoxyhemoglobin along with the blood flow, volume and oxygenation change in response to the neural activity. Figure 2.2 illustrates the hemodynamic response corresponding to the neural activation after a stimulus onset. When the neurons in a brain area are activated, they require more glucose and oxygen, and thus an increase of local cerebral blood flow and blood volume occurs in order to satisfy the excess demand of energy consumption in this active brain region. After a delay of around 1–5 seconds, a more rapid oxygen release can normally be seen in this active area than the rest areas with a initial dip, which is followed by an overshoot of the MRI signal. This process is called hemodynamic response, which leads to the local raise and drop in the relative concentration of oxyhemoglobin and deoxyhemoglobin in veins respectively during activation even if the active neurons consume more oxygen by altering oxyhemoglobin to deoxyhemoglobin. The magnetic properties of oxyhemoglobin and deoxyhemoglobin vary since the oxyhemoglobin is diamagnetic and the deoxyhemoglobin is paramagnetic, and the difference in magnetic susceptibility between oxyhemoglobin and deoxyhemoglobin gives rise to the magnetic signal variation for MRI detection and allows the spatial localization and functional mapping of the activated brain region. After the activation, the received signal falls immediately and returns to the original level or the baseline, which is generally fol-

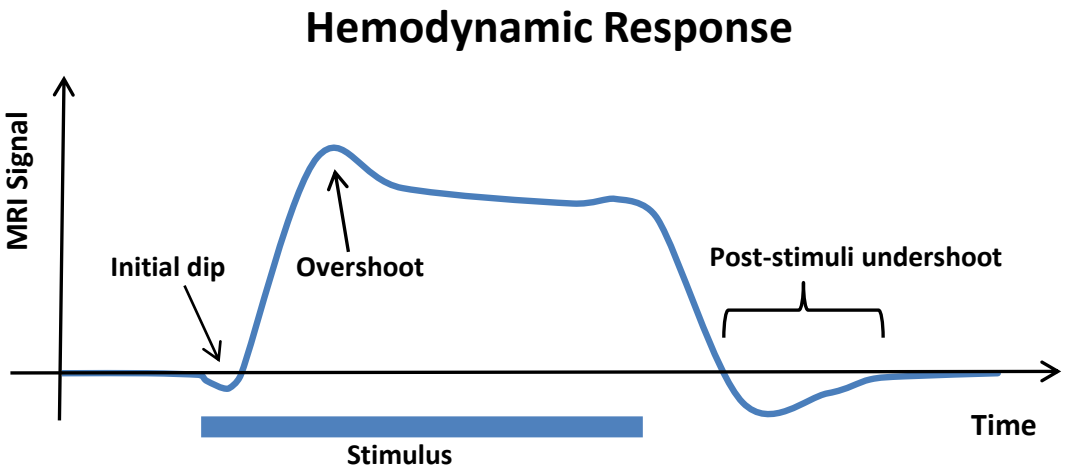


Figure 2.2: With the stimulus onset, the hemodynamic response corresponding to neural activity starts with a initial dip and then is followed by an overshoot of the MRI signal. After neural activation, a post-stimuli undershoot below the baseline can be seen after a rapid decline in MRI signal.

lowed by a post-stimuli undershoot below the baseline. In 1990, Ogawa et al. (1990) firstly described that the changes in blood flow oxygenation level corresponding to the neural activation can be used for the functional brain image generation. This imaging approach is called BOLD contrast, which is sensitive to the effects of hemodynamic response and has been widely applied in many fields. The predicted BOLD time series, which include a large number of images acquired in a temporal order, can be modeled and obtained by convolving the predicted neural activity curve based on the experimental design with a canonical hemodynamic response function (HRF), where a good match between the predicted and observed response implies the activation is related to the task or stimuli.

2.2 Pre-processing of fMRI Data

The pre-processing procedure, an essential part of neuroimaging data analysis, generally includes an ordered sequence of individual image and signal processing steps, which varies between different software packages. The raw fMRI BOLD time series from the MRI scanner are quite variable with fairly low functional signal-to-noise ratio (SNR) and require pre-processing to prepare the right data prior to statistical analysis. Pre-processing attempts to reduce the noise within the data caused by ab-

normalities and artifacts from the scanner or experimental subject and to correct for the variability of no interest from the data unrelated with the task. Pre-processing also assists us in preparing the data to meet the assumptions needed in the later analysis (Poldrack et al., 2011).

Although the subjects are requested to avoid moving their head during the experiment, in practice some involuntary head motion is unavoidable. Even small head motions can corrupt the activation-related BOLD signal changes, and thus the pre-processing step of image realignment is requisite for motion correction by realigning each fMRI volume to a selected common reference. In addition, there is a considerable variation between individuals in brain size and shape. The pre-processing step known as spatial normalization consists of warping each individual's brain into a common reference space, and allowing for the examination of fMRI BOLD signal changes across individuals within a group or between groups. One commonly used reference brain space is the Talairach atlas based on a single subject, where certain landmarks are used to transform each brain into a common pose and size; another is the MNI atlas, based on a population of subjects. Other artifacts and distortions induced by various reasons, such as spin-history changes, RF interference, inhomogenous magnetic field, physiological cardiac and respiratory pulsations, and instrumental instability, can be similarly dealt with by the corresponding pre-processing methods (e.g., slice timing correction and image reconstruction). The final pre-processing step is usually spatial smoothing. The commonly used method of Gaussian smoothing blurs the images with a Gaussian kernel function for the purpose of removing the isolated points of noise in the image with improved SNR, and compensating the imperfect spatial normalization. Another motivation for smoothing is to improve the Gaussian behaviour, due to the averaging of the convolution.

2.3 Statistical Modeling of fMRI Data

After obtaining the pre-processed fMRI BOLD time series, the statistical analysis proceeds with the intra-subject modeling, which assesses the single-subject data for the evidence against the null hypothesis of no task- or stimulus-related effects, and then generates the statistic output image for each subject. This intra-subject model can also be embedded in the group model to form a single model investigating all subjects simultaneously for longitudinal data analysis, but normally, it would be more intuitive and more efficient to separate this single statistical model into two parts: first level or intra-subject modeling, and group level or between-subject mod-

eling, and extract the information needed from the first level to yield the group data.

For each voxel, the first-level or intra-subject modeling is to build and fit the general linear model (GLM) to a single subject's fMRI BOLD time series with one scan per timepoint (Frackowiak et al., 2004). Assume that the sample size is n. The first-level GLM of subject i $(i = 1, \ldots, n)$ for a particular voxel can be described in a matrix form as follows:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \ \boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}, \mathbf{V}_i),$$

where $\mathbf{Y}_i$ is an $s_i \times 1$ vector of BOLD time series, $\mathbf{X}_i$ is an $s_i \times p_i$ design matrix with the use of the predicted hemodynamic response to each task or stimulus as a regressor, $\boldsymbol{\beta}_i$ is the unknown time-invariant $p_i \times 1$ regression parameter vector (or effect sizes for all tasks or stimuli), $\boldsymbol{\epsilon}_i$ is the within-subject random error, and $\mathbf{V}_i$ is the within-subject covariance matrix (Beckmann et al., 2003; Frackowiak et al., 2004; Muford and Nichols, 2006). Here, s_i denotes the number of scans for subject i, which can vary between subjects, and p_i specifies the number of explanatory variables related to the tasks or stimuli according to the experimental design. Denote the overall voxel-wise output statistic values for the first-level inference as a vector $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_n)^{\mathrm{T}}$ from these n subjects after ignoring the voxel index. Here $\boldsymbol{\varphi}$ consists of all the information of interest extracted from the first-level inference and its element $\boldsymbol{\varphi}_i$ $(i = 1, \ldots, n)$ can be true values of the parameter vector $(\boldsymbol{\varphi}_i = \boldsymbol{\beta}_i)$, estimates of the parameter vector $(\boldsymbol{\varphi}_i = \widehat{\boldsymbol{\beta}}_i)$, contrast $(\varphi_i = \mathbf{c}^{\mathrm{T}}\boldsymbol{\beta}_i)$, or contrast estimate $(\varphi_i = \mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_i)$, where the contrast coefficient vector $\mathbf{c}$ is pre-defined according to the experimental requirement. The matrix form of group-level or between-subject GLM for a specified voxel is

$$\boldsymbol{\varphi} = \mathbf{G}\boldsymbol{\theta} + \boldsymbol{\eta}, \ \boldsymbol{\eta} \sim \mathrm{N}(\mathbf{0}, \mathbf{M}),$$

where $\mathbf{G}$ is the $(\sum_{i=1}^{n} p_i) \times p$ group-level design matrix, involving subject-specific covariates (e.g., age and gender), $\boldsymbol{\theta}$ is a p-vector of the p unknown between-subject regression coefficients, $\boldsymbol{\eta}$ is a $(\sum_{i=1}^{n} p_i) \times 1$ vector specifying the between-subject random error, and $\mathbf{M}$ represents the between-subject covariance matrix.

If the summarized statistic of the information from first-level inference is simply the true parameter values $\boldsymbol{\beta}_i$ $(i = 1, \ldots, n)$ for subject i, then the group data $\boldsymbol{\varphi}$ can be written as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_n)^{\mathrm{T}}$, and the integration of the intra-subject and

between-subject models provides a single model below:

$$\mathbf{Y} = \mathbf{X}\mathbf{G}\boldsymbol{\theta} + \boldsymbol{\gamma},\ \boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\eta} + \boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \mathbf{X}\mathbf{M}\mathbf{X}^{\mathrm{T}} + \mathbf{V}), \tag{2.2}$$

where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix},\quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_n \end{pmatrix},\quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_n \end{pmatrix},$$

and the covariance matrix $\mathbf{V}$ is a block-diagonal matrix with the block $\mathbf{V}_i$, as previously mentioned, specified as the within-subject covariance matrix for subject i (Beckmann et al., 2003).

As the true within-subject parameter values (effect sizes) are unknown in practice, the group data $\boldsymbol{\varphi}$ can be derived as the estimated parameter values from first-level inference, the contrast estimate (defined as a linear combination of the parameter estimates), or the contrast-related test statistic value using t-test or F-test, e.g., $\varphi_i = \mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_i$ for some p_i-vector $\widehat{\boldsymbol{\beta}}_i$. The formation of the group data is based on the experimental protocol, and the structure of between-subject covariance matrix can be distinct for different sample formations. For example, when we consider the single model (2.2), it can be equally separated into two levels and the group data is the unknown first-level parameter vector $\boldsymbol{\beta}$. We consider adapting the generalized least squares parameter estimate $\widehat{\boldsymbol{\beta}}$ as the group data instead of the unknown parameter vector. In this case, the combination of the separated intra-subject and between-subject models can be equivalent to the single model (2.2) in respect of parameter estimation only if the group-level GLM is constructed as

$$\widehat{\boldsymbol{\beta}} = \mathbf{G}\boldsymbol{\theta} + \widetilde{\boldsymbol{\eta}},\ \widetilde{\boldsymbol{\eta}} \sim \mathrm{N}(\mathbf{0}, \widetilde{\mathbf{M}}),$$

where $\widetilde{\mathbf{M}} = \mathbf{M} + (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}$ (Beckmann et al., 2003). When the task-related or stimulus-response brain activation is investigated, different types of events including various tasks (or stimuli) and the rest baseline are sequentially designed, and the observed single subject's data is compared against our prediction for the whole time series based on the convolution of HRF and neural event timings (activation patterns). If the within-subject contrast estimates from all subjects are treated as the group data and written in a vector notation as

$$\boldsymbol{\varphi} = \left(\mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_1, \mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_2, \ldots, \mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n\right)^{\mathrm{T}},$$

$\boldsymbol{\varphi}$ is also normally distributed as each of its element $\varphi_i = \mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_i$ $(i = 1, \ldots, n)$ follows a Gaussian distribution, where the variance for φ_i is

$$\mathbb{V}\mathrm{ar}(\varphi_i) = \mathbf{c}^{\mathrm{T}}\mathbb{C}\mathrm{ov}(\widehat{\boldsymbol{\beta}}_i)\mathbf{c},$$

and then the covariance matrix of $\boldsymbol{\varphi}$ can be correspondingly constructed with this equation (Frackowiak et al., 2004).

2.4 Heritability and Twin Studies

Heritability can be interpreted as the proportion of phenotypic variation explained by a single genetic marker (Filippini et al., 2008) or the entire genome (Stein et al., 2010). In quantitative genetics, the phenotypic variance of an observable trait, denoted as σ^2_{P}, can be seen as the sum of independent genetic (σ^2_{G}) and environmental (σ^2_{E}) contributions (Falconer and Mackay, 1996), as per

$$\sigma^2_{\mathrm{P}} = \sigma^2_{\mathrm{G}} + \sigma^2_{\mathrm{E}}.$$

The heritability in the broad sense measures the overall genetic influence σ^2_{G} on a phenotypic trait:

$$H^2 = \frac{\sigma^2_{\mathrm{G}}}{\sigma^2_{\mathrm{P}}},$$

where H^2 represents the broad-sense heritability.

The total genetic variation is comprised of additive and non-additive genetic factors. The additive genetic factor is influenced by the linear addition of the independent genes, or more technically, allelic contributions at different gene loci, while the non-additive genetic effect refers to the influence of the interactions among alleles within or between gene loci, such as dominance and epistasis effects. The heritability in the narrow sense is defined as the proportion of phenotypic variation accounted for by the additive genetic effect (σ^2_{A}):

$$h^2 = \frac{\sigma^2_{\mathrm{A}}}{\sigma^2_{\mathrm{P}}},$$

where h^2 denotes the narrow-sense heritability.

The additive genetic variation is usually of more interest because the additive factor is due to the sum of mean effects of the alleles at a particular locus or at multiple

trait-related loci, and only this additive genetic effect can be passed down from one generation to the next. Hence we will follow the typical convention, and use the term "heritability" to refer to narrow-sense heritability unless otherwise noted. Heritability of a particular trait can vary between 0 and 1, where a heritability of 0 means that there are no additive genetic effects at all, while a heritability of 1 implies the additive genetic factor completely explains this trait.

There are generally two types of twins, identical or monozygotic (MZ) twins and fraternal or dizygotic (DZ) twins. MZ twins have exactly identical genes and DZ twins share 50% genes on average, which leads to the sharing of additive genetic effects. Even in the absence of a genetic influence on the response, it is possible that twins are similar since each has been raised in the same family environment. This gives rise to the common environmental factor, which induces the covariance within twin pairs regardless of MZ or DZ type. Finally there is an independent unique error, corresponding to the usual independent and identically distributed (i.i.d.) noise corrupting the measurements.

Therefore the phenotypic variance within a population is assumed to be homogeneous and can be partitioned into three components:

$$\sigma^2 = A + C + E,$$

where $\sigma^2 = \sigma^2_{\mathrm{P}}$ denotes the phenotypic variation (subsequently, we suppress the subscript "P"), and A, C and E represent the additive genetic, common environmental and unique environmental components of the phenotypic variance. The so-called ACE modeling in twin studies is based on this variance decomposition (Lee et al., 2010). In the ACE model, then, the narrow-sense heritability is

$$h^2 = \frac{A}{A + C + E}, \tag{2.3}$$

and the modeling of heritability consists of the modelling of these three variance components.

2.5 Inference on Heritability

This section concentrates on the introduction of the commonly used terminology on heritability inference, the permutation approach for constructing the empirical distribution of test statistic and computing the permutation-based p-value, and

the bootstrap procedure for calculating bootstrapping confidence intervals for the statistic or parameter of interest.

2.5.1 Terminology on Heritability Analysis

Hypothesis testing is a fundamental statistical tool. It is based on some statement of the population under a default state, referred to as the null hypothesis and denoted as H_0; the test then measures the evidence against this hypothesis, allowing one to (possibly) reject H_0 in favor of an alternate hypothesis H_1, the negation of H_0. The evidence against H_0 is measured with a test statistic, which is a numerical summary extracting the interested information from the sample for comparison purposes, in order to distinguish the null hypothesis in favor of the alternative hypothesis. For a single test, the result of rejecting the null hypothesis at a pre-determined significance level α (e.g., $\alpha = 0.05$), which gives rise to the critical value that is a threshold for the test statistic deciding the rejection of the null hypothesis, is called statistically significant and is unlikely to have happened by chance. The rejection of the null hypothesis at a given level α can also be determined by the p-value, which is defined as the conditional probability of obtaining the test statistic at least as extreme as the observed value under the null hypothesis, and the result of p-value less than the significance level indicates that the observed result is not arisen by chance and the null hypothesis should be rejected.

Conventionally, the sampling distribution of the test statistic assuming the null hypothesis is true, which is called null distribution, is exactly or approximately computable. When the null distribution is not tractable or accurate, permutation inference can be utilised (described below). When testing a null hypothesis, we may arrive at a wrong decision if the result does not correspond with the reality. If the true null hypothesis is incorrectly rejected at a given level α, i.e., the value of the test statistic lies within the rejection region: $[T_\alpha, \infty]$ for a one-tailed test or $[-\infty, T_{\frac{1}{2}\alpha}] \cup [T_{1-\frac{1}{2}\alpha}, \infty]$ for a two-tailed test, where T_α is the α-level critical value, or the corresponding p-value is smaller than α while the null hypothesis is actually true, we call a type I error or false positive occurs; if we fail to reject the false null hypothesis, i.e., the value of the test statistic is outside the rejection region or the corresponding p-value exceeds the level α while the null hypothesis is invalid, we call a type II error or false negative occurs. The occurrence of both type I and type II errors should be controlled, but there is a trade-off between the probability of making these two types of errors (type I and type II error rates) since reducing one type of error generally induces an increase in another for the same sample size.

Type I error rate, also called the false positive rate (FPR), relates to the proportion of correct identification of true negatives (or specificity) of a test and type II error rate is related to the statistical power (or sensitivity) of the test. The only way to reduce both error rates without improving the test is to increase the sample size; while for a given sample, one of these error rates is constrained by a given level α, and another is minimized by improving the sensitivity or specificity of the test.

Until now, we only focus on assessing a single particular test, and now we will consider testing multiple hypotheses simultaneously. When there are multiple, say m, comparisons in total to be performed, there is no longer a unique measure of the false positive risk. The most commonly used measure, the family-wise error rate (FWER) is defined as the probability of any false positives occurring among the m tests. If all the tests were independent, and a significance level of α was used for each test, the FWER would be controlled at level α_{FWER},

$$\alpha_{\text{FWER}} = 1 - (1 - \alpha)^m.$$

Another measure, the false discovery rate (FDR) is the expected proportion of false discoveries among all discoveries (i.e., the expected proportion of false positives among all rejections of the null hypothesis), resulting in a higher statistical sensitivity for the FDR control than the FWE procedure at the expense of increased FPR. The procedure of FDR control works by correcting and adjusting the p-value for each test, which was introduced explicitly by Benjamini and Hochberg (Benjamini and Hochberg, 1995). For those m tests, denote the derived p-values in an ascending order as

$$p_{(1)} \leq \cdots \leq p_{(m)},$$

and denote the corresponding null hypotheses as $\mathrm{H}_0^{(i)}$ $(i = 1, \ldots, m)$. If $i_{\max}$ is the largest index i satisfying

$$p_{(i)} \leq \frac{i}{mc(m)}\alpha,$$

where α is a fixed threshold within $[0, 1]$ and $c(m) = 1$ or $c(m) = \sum_{i=1}^{m} \frac{1}{i}$, then all hypotheses $\mathrm{H}_0^{(i)}$ $(i = 1, \ldots, i_{\max})$ should be rejected at level α, and the FDR-corrected p-values are defined and calculated as

$$q_{(i)} = \frac{p_{(i)}m}{i}c(m).$$

The choice of $c(m)$ is based on the correlation assumption for the tests: $c(m) = 1$ is more sensitive and valid when independence or positive dependence is assumed;

$c(m) = \sum_{i=1}^{m} \frac{1}{i}$ is valid under any situations (Benjamini and Yekutieli, 2001). As $q_{(i)}$ is not a monotonic function of $p_{(i)}$, the FDR-adjusted p-values are further defined as

$$\widetilde{q}_{(i)} = \min\Big(q_{(i)}, \ldots, q_{(m)}\Big)$$

to ensure the monotonicity (Yekutieli and Benjamini, 1999). For the imaging data, the assumption of positive dependence is reasonable, and thus $c(m) = 1$ should be chosen with more sensitivity.

2.5.2 Permutation Test and FWE Correction

Permutation test is a non-parametric technique that makes minimal assumptions about the data. With few simple assumptions like exchangeability of the observed data under the null hypothesis, the non-parametric permutation test is conceptually simple and theoretically intuitive (Nichols and Hayasaka, 2003; Nichols and Holmes, 2001). When the null hypothesis is true, the data will exhibit a form of exchangeability, allowing permutation, re-fitting the model and computation of test statistic. With multiple permutations, an empirical null distribution can be constructed and critical thresholds and p-values computed. Currently this approach has become well-known and widespread gradually owing to the recent development of inexpensive and powerful computers.

Applying variance component inference approach voxel-by-voxel yields a test statistic image. For each voxel, if the null hypothesis of no heritability, $\mathrm{H}_0 : h^2 = 0$, is assumed to be true, the MZ and DZ twin pairs become exchangeable, allowing $\mathrm{P} = \binom{(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})/2}{n_{\mathrm{MZ}}/2}$ possible permutations in total. Normally a small-to-medium sample gives a comparatively large P, so an approximate, or Monte Carlo permutation test can be exploited instead with a smaller number of permutations, say N, based on a random subsample of all permutations (Nichols and Holmes, 2001).

In order to resolve the multiple comparisons problem and strictly control the false positives over the whole volume of the ROI's simultaneously, permutation test can be employed to implement the procedure of FWE correction. Type I errors for all these comparisons are under strong control, and FWE-corrected p-values are computed by considering the maximum test statistics (Nichols and Holmes, 2001). With a permutation test, we obtain FWE-corrected p-values on peak height (voxel-wise test statistic value) for voxel-wise inference, and cluster size (number of voxels involved in a cluster after thresholding) and cluster mass (sum of voxel-wise test

statistic values from all voxels within a cluster after thresholding) for cluster-level inference. Hence, this permutation approach can be further separated into two parts: voxel-wise single threshold test and cluster-wise suprathreshold tests.

Voxel-wise Single Threshold Test

Along with the single threshold test, the omnibus null hypothesis of no heritability for all voxels over the ROI's is rejected if any voxel-wise test statistic value exceeds a given critical threshold, or equivalently, if the maximum voxel-wise test statistic value exceeds this threshold. The critical threshold is pre-defined with the significance level α. By permuting the labels of MZ and DZ for twin pairs and computing the test statistic image for each permutation, the empirical distribution of this maximum test statistic can be constructed using the maximum test statistic values obtained from all permutations, and the critical threshold is the $c+1$ largest value of the empirical distribution, where $c = \lfloor \alpha \mathrm{N} \rfloor$ (Nichols and Holmes, 2001).

The procedure of single threshold test is detailed as follows. For permutation p ($\mathrm{p} = 1, \ldots, \mathrm{N}$), the maximum test statistic over the ROI's is denoted as $\mathrm{T}_{\mathrm{p}}^{\max}$. The original (unpermuted) data and the other $\mathrm{N}-1$ relabelings together provide totally N maximum test statistic values, which can be used to create the empirical distribution of this maximum test statistic, and then provide the α-level critical threshold as the $c + 1 = \lfloor \alpha \mathrm{N} \rfloor + 1$ largest value, denoted as T_α. If the unpermuted maximum test statistic value is greater than T_α, the omnibus hypothesis can be rejected. The corresponding FWE-corrected p-value, written as p_{FWE}, for each voxel can be calculated as the proportion of these N maximum test statistic values in the constructed permutation-based empirical distribution that are not smaller than the original voxel-wise test statistic value (T_0):

$$p_{\mathrm{FWE}} = \frac{\#\{\mathrm{T}_{\mathrm{p}}^{\max} \geq \mathrm{T}_0\}}{\mathrm{N}}.$$

Cluster-wise Suprathreshold Tests

The significance of suprathreshold cluster tests can be assessed by the spatially informed cluster statistics, such as cluster size and cluster mass. A pre-selected cluster-forming threshold, which can be expressed as a p-value using the sampling distribution of the test statistic, is given and applied to the derived test statistic image to threshold test statistic values and form suprathreshold clusters, which are brain regions of connecting voxels with the test statistic values above that cluster-

forming threshold. Accordingly, the suprathreshold cluster size and suprathreshold cluster mass are defined as the number of voxels in a suprathreshold cluster and the sum of voxel-wise test statistic values within a suprathreshold cluster respectively. Theoretically similar to the single threshold test, suprathreshold cluster tests require constructing the empirical distribution of the maximum suprathreshold cluster statistics with permutations. The pre-determined significance level α also provides the critical threshold to be the $c + 1 = \lfloor \alpha \mathrm{N} \rfloor + 1$ largest member within the empirical distribution.

The mechanics of the cluster-wise suprathreshold tests are described as follows. For permutation p ($\mathrm{p} = 1, \ldots, \mathrm{N}$), the maximum suprathreshold cluster size (or cluster mass) is denoted as $\mathrm{K}_\mathrm{p}^{\max}$ (or $\mathrm{M}_\mathrm{p}^{\max}$). An original data and the other $\mathrm{N} - 1$ relabelings (permuted data vectors) are analyzed, and the resulting N measures of maximum suprathreshold cluster size (or cluster mass) are sorted to form the empirical distribution of this cluster statistic. The critical threshold at level α can be calculated as the $c + 1 = \lfloor \alpha \mathrm{N} \rfloor + 1$ largest member of the empirical distribution, denoted as K_α (or M_α). Significance of the test is determined by whether or not the original maximum suprathreshold cluster size (or cluster mass) is greater than K_α (or M_α). The associated FWE-corrected p-value, denoted as p_{FWE}, for each individual suprathreshold cluster on the original test statistic image can be computed as the proportion of these N measures of maximum suprathreshold cluster size (or cluster mass) within the empirical distribution greater than or equal to the observed size (or mass) of that cluster (Nichols and Holmes, 2001), i.e.,

$$p_{\mathrm{FWE}} = \frac{\#\{\mathrm{K}_\mathrm{p}^{\max} \geq \mathrm{K}_0\}}{\mathrm{N}},$$

$$p_{\mathrm{FWE}} = \frac{\#\{\mathrm{M}_\mathrm{p}^{\max} \geq \mathrm{M}_0\}}{\mathrm{N}},$$

for cluster statistics of size and mass respectively.

2.5.3 Bootstrapping Confidence Interval

While permutation is a straightforward procedure to obtain p-values, it is difficult to obtain confidence intervals. Hence we use bootstrapping technique for constructing confidence intervals. The bootstrapping confidence intervals for unknown population parameters are conceptually simple and based on a simple idea of resampling the data with replacement from the observed sample. For each original sample, the model is re-fitted to each bootstrap sample, and the sampling distribution is ap-

proximated by its bootstrap resampling distribution (for a sufficiently large number of bootstrap replicates, e.g., B = 1000). A so-called balanced bootstrap algorithm is generally preferable and can be performed to supply the required B bootstrap replicates, where each observation from the original sample is equally used B times in all bootstrap samples (Gleason, 1988).

For twin studies, the bootstrap resampling is stratified by separating the original sample data into a MZ group, a DZ group and a singleton group, and then drawing bootstrap samples individually from these three groups with the same size and structure as the original sample; specifically, twins are always sampled in pairs within MZ or DZ group. For each bootstrap sample, the evaluated statistic is denoted as T^{b}. The empirical bootstrap distribution of the statistic of interest is used to compute the confidence intervals via the standard error or percentiles (DiCiccio and Efron, 1996).

In this section, we will propose a joint method of standard bootstrap, percentile bootstrap and bias-corrected percentile bootstrap for the construction of bootstrapping confidence intervals (CIs).

The standard bootstrap $100(1-\alpha)\%$ CI is defined as

$$\Big[\mathrm{T}_0 - z_{1-\frac{1}{2}\alpha}\mathrm{s}^{\mathrm{b}},\ \mathrm{T}_0 + z_{1-\frac{1}{2}\alpha}\mathrm{s}^{\mathrm{b}}\Big],$$

where T_0 is the observed value of the statistic from the original data, $z_{1-\frac{1}{2}\alpha}$ is the $100(1-\frac{1}{2}\alpha)\%$ percentile of the standard normal distribution, and s^{b} is the bootstrap sample standard deviation. The standard bootstrap does not perform well when the data is highly non-normal. The percentile bootstrap CI uses the empirical percentiles of the bootstrap distribution with the $100(1-\alpha)\%$ interval of

$$\Big[\mathrm{T}^{\mathrm{b}}_{\frac{1}{2}\alpha},\ \mathrm{T}^{\mathrm{b}}_{1-\frac{1}{2}\alpha}\Big],$$

where $\mathrm{T}^{\mathrm{b}}_{\frac{1}{2}\alpha}$ and $\mathrm{T}^{\mathrm{b}}_{1-\frac{1}{2}\alpha}$ denote the $100(\frac{1}{2}\alpha)\%$ and $100(1-\frac{1}{2}\alpha)\%$ percentiles of the bootstrap distribution respectively. While the percentile CI is intuitive, it too can have poor coverage. The bias-corrected percentile bootstrap CI adjusts the bias in bootstrap distribution for a better approximation. The bias-adjusted $100(1-\alpha)\%$ interval is:

$$\Big[\Phi(2z_0 - z_{1-\frac{1}{2}\alpha}),\ \Phi(2z_0 + z_{1-\frac{1}{2}\alpha})\Big],$$

where

$$z_0 = \Phi^{-1}\Big[\frac{\#\{\mathrm{T}^{\mathrm{b}} \leq \mathrm{T}_0\}}{\mathrm{B}+1}\Big],$$

and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

We found that none of these 3 methods provided accurate coverage in the context of heritability inference, so we created a joint method based on the union of intervals from these methods. Denote the interval using our joint method as $[\mathrm{T}^{\mathrm{b}}_{\mathrm{L}},\ \mathrm{T}^{\mathrm{b}}_{\mathrm{U}}]$. The lower and upper bounds of this interval are characterized as the minimum of the lower bounds of the above three intervals and the maximum of the upper bounds of those intervals respectively:

$$\mathrm{T}^{\mathrm{b}}_{\mathrm{L}} = \min\Big(\mathrm{T}_0 - z_{1-\frac{1}{2}\alpha}\mathrm{s}^{\mathrm{b}},\ \mathrm{T}^{\mathrm{b}}_{\frac{1}{2}\alpha},\ \Phi(2z_0 - z_{1-\frac{1}{2}\alpha})\Big),$$

$$\mathrm{T}^{\mathrm{b}}_{\mathrm{U}} = \max\Big(\mathrm{T}_0 + z_{1-\frac{1}{2}\alpha}\mathrm{s}^{\mathrm{b}},\ \mathrm{T}^{\mathrm{b}}_{1-\frac{1}{2}\alpha},\ \Phi(2z_0 + z_{1-\frac{1}{2}\alpha})\Big).$$

In simulation evaluations we found this joint bootstrap method has a favorable performance, generating intervals with accurate coverage probability and reasonable length.

There exists a close correspondence between hypothesis tests and confidence intervals. If the plausible value of the statistic of interest, specified by the null hypothesis, is not included within the constructed confidence interval at a given level α, it supplies an evidence, provided by the confidence interval, against the null hypothesis at level α. Thus, the confidence interval can also be treated as a complement of hypothesis testing aiding in the assessment of the null hypothesis.

2.5.4 Summary Statistics

The test statistic for hypothesis testing is a quantity derived from the sample in order to measure the compatibility between the null hypothesis and the sample data, and determine whether this null hypothesis should be rejected or not. Test statistics developed from a likelihood ratio are optimally powerful according to the Neyman-Pearson lemma, under certain conditions. Other types of test statistics, however, may also be useful even if not theoretically optimal. A statistic that is interpretable and captures the differences between the observed data and the null-hypothesized models may indeed be useful.

Conventionally, hypothesis testing utilizes test statistics whose exact or approximate

theoretical null distribution is known under certain strong assumptions of the data such as normality. The permutation test, nevertheless, has an important property of allowing the use of non-standard test statistics with unknown or complicated null distribution (Winkler et al., 2014). Owing to this key feature of permutation test, we also considered employing the useful whole-image summary statistics as test statistics in addition to cluster statistics for the analysis of the imaging data: the unweighted ($\overline{h^2}$) and variance-weighted ($\overline{wh^2}$) averages of all voxel-wise heritability estimates, the second (Q_2, the median) and third (Q_3) quartiles of these estimates, mean of those heritability estimates greater than Q_2 ($\overline{h^2}(Q_2)$), and mean of those heritability estimates greater than Q_3 ($\overline{h^2}(Q_3)$). These statistics emphasize the right "tail" (e.g., we omit the first quartile Q_1), as exact-zero h^2 values make interpreting the left "tail" difficult. If we assume that there are totally K in-mask voxels within the ROI's, these mean statistics are defined as follows:

$$\overline{h^2} = \frac{1}{K}\sum_{r=1}^{K} h_r^2,$$

$$\overline{wh^2} = \frac{1}{K}\sum_{r=1}^{K}\left(\sigma_r^2 \Big/ \overline{\sigma^2}\right) h_r^2, \quad \overline{\sigma^2} = \frac{1}{K}\sum_{r=1}^{K}\sigma_r^2,$$

$$\overline{h^2}(Q_2) = \frac{\#\{h_r^2 > Q_2\}}{K},$$

$$\overline{h^2}(Q_3) = \frac{\#\{h_r^2 > Q_3\}}{K},$$

where σ_r^2 and h_r^2 denote the voxel-wise phenotypic variance and the corresponding heritability for voxel r ($r = 1, \ldots, K$).

The permutation inference can be implemented rapidly using these summary statistics, their empirical null distribution can be formed by permutation test, and the corresponding permutation-based p-values can be obtained using these null distributions. The fast implementation of these summary statistics provides a tool for exploring the whole brain quickly and a significant result with p-values less than the given level α implies that there should be some significantly heritable brain regions. Moreover, the utilization of bootstrap procedure can provide the bootstrapping confidence intervals for these summary statistics, which can also act as the complement to assessing the corresponding hypothesis test.

Chapter 3

Mass-univariate Heritability Inference

This chapter begins with the construction of an univariate model describing the covariance structure of the data at each voxel. At each voxel, the variance component parameters are estimated, and then heritability estimates obtained. We briefly review those existing heritability estimation methods, and then discuss our newly proposed methods for heritability estimation in detail. In this chapter we only consider a sample comprised of twins and singletons (individuals unrelated to any other subjects).

3.1 The General Linear Model

Suppose that there are $\frac{1}{2}n_{\mathrm{MZ}} > 0$ MZ twin pairs (n_{MZ} individuals), $\frac{1}{2}n_{\mathrm{DZ}} > 0$ DZ twin pairs (n_{DZ} individuals), and $n_{\mathrm{S}} \geq 0$ singletons (unrelated subjects), and in total $n = n_{\mathrm{MZ}} + n_{\mathrm{DZ}} + n_{\mathrm{S}}$ participants in the experiment. For a particular voxel $r \in \{1, \ldots, K\}$, the column vector $\mathbf{Y}_r^{(k)}$ is used to denote the observations from all these n participants for the phenotype $k \in \{1, \ldots, J\}$:

$$\mathbf{Y}_r^{(k)} = \Big(\begin{array}{lll} \mathrm{Y}^{(k)}_{1,1,r} & , \ \mathrm{Y}^{(k)}_{2,1,r} & , \ \cdots \ , \\ \mathrm{Y}^{(k)}_{n_{\mathrm{MZ}}+1,\frac{1}{2}n_{\mathrm{MZ}}+1,r} & , \ \mathrm{Y}^{(k)}_{n_{\mathrm{MZ}}+2,\frac{1}{2}n_{\mathrm{MZ}}+1,r} & , \ \cdots \ , \\ \mathrm{Y}^{(k)}_{n_{\mathrm{MZ}}+n_{\mathrm{DZ}}+1,0,r} & , \ \mathrm{Y}^{(k)}_{n_{\mathrm{MZ}}+n_{\mathrm{DZ}}+2,0,r} & , \ \cdots \end{array} \Big)^{\mathrm{T}}.$$

That is, $\mathrm{Y}^{(k)}_{i,j,r}$ represents the observation for voxel r from subject i (twin or singleton, $i = 1, \ldots, n$), where j ($j = 0, 1, \ldots, \frac{1}{2}(n_{\mathrm{MZ}} + n_{\mathrm{DZ}})$) is the index of twin pair with $j = 0$ for singletons. While somewhat cumbersome, we need this detailed notation

for later expressions. Note that we always arrange phenotypic data elements so that twin pairs are adjacent, and so that all MZ twins come first, then DZ twins and then finally singletons.

Some types of brain imaging data are directly measured, like gray matter density, producing one image per subject. However fMRI, in distinction, is a derived measure based on hundreds of scans per subject to estimate blood flow change. As reviewed in Section 2.3, for each subject, an intra-subject model is fitted to the time series for each of the voxels, producing an image of BOLD effect magnitude (Frackowiak et al., 2004). If an fMRI phenotype is used we presume that a BOLD effect contrast has been computed, one measure per subject, per voxel. Thus, for voxel r, the group GLM adopted in genetic variance components analysis for phenotype k is constructed in a matrix form as

$$\mathbf{Y}^{(k)} = \mathbf{X}^{(k)}\boldsymbol{\beta}^{(k)} + \boldsymbol{\epsilon}^{(k)}, \tag{3.1}$$

where $\mathbf{X}^{(k)}$ is the $n \times p$ design matrix including an intercept and $p-1$ covariates (Typical covariates would include age, gender, or other inter-subject effects.), $\boldsymbol{\epsilon}^{(k)}$ is the error vector assumed to be normally distributed, written as $\boldsymbol{\epsilon}^{(k)} \sim \mathbb{N}(\mathbf{0}, \mathbf{V}^{(k)})$, and we have suppressed the subscript r, as we will fit the same form of model at each voxel; the definition of the variance-covariance matrix $\mathbf{V}^{(k)}$ is detailed below. In this chapter, the univariate case is considered, therefore the superscript index k for varying phenotypes can also be suppressed.

To simplify the description of the variance components, we introduce a subject type indicator function $\mathcal{T} : \{1, \ldots, n\} \to \{\mathrm{MZ}, \mathrm{DZ}, \mathrm{S}\}$ that maps a subject index i to its corresponding subject type MZ, DZ or S. For the data arrangement we prescribe, this mapping is

$$\mathcal{T}(i) = \begin{cases} \mathrm{MZ}, & \text{if } i = 1, \ldots, n_{\mathrm{MZ}}, \\ \mathrm{DZ}, & \text{if } i = n_{\mathrm{MZ}} + 1, \ldots, n_{\mathrm{MZ}} + n_{\mathrm{DZ}}, \\ \mathrm{S}, & \text{otherwise.} \end{cases}$$

The ACE model (see Section 2.4) gives the covariance for MZ twin pair j ($j =$

$1, \ldots, \frac{1}{2}n_{\mathrm{MZ}}$) as

$$\begin{aligned}\mathbb{C}\mathrm{ovMZ} &= \mathbb{C}\mathrm{ov}\begin{pmatrix}\mathrm{Y}_{2j-1,j}\\ \mathrm{Y}_{2j,j}\end{pmatrix}\\ &= \begin{pmatrix}A+C+E & A+C\\ A+C & A+C+E\end{pmatrix},\end{aligned} \tag{3.2}$$

where A, C and E denote the additive genetic, common environmental, and unique error components, respectively. Correspondingly, the covariance of DZ twin pair j ($j = \frac{1}{2}n_{\mathrm{MZ}}+1, \ldots, \frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})$) is

$$\begin{aligned}\mathbb{C}\mathrm{ovDZ} &= \mathbb{C}\mathrm{ov}\begin{pmatrix}\mathrm{Y}_{2j-1,j}\\ \mathrm{Y}_{2j,j}\end{pmatrix}\\ &= \begin{pmatrix}A+C+E & \frac{1}{2}A+C\\ \frac{1}{2}A+C & A+C+E\end{pmatrix}.\end{aligned} \tag{3.3}$$

Note how $\mathbb{C}$ovMZ (3.2) has a covariance of $A+C$, whereas $\mathbb{C}$ovDZ (3.3) has a covariance of $\frac{1}{2}A+C$, reflecting how DZ twin pairs share only half of their genetic material; both have the same term C, as a common environment equally affects MZ and DZ twin pairs. Each singleton is unrelated with all other subjects, and therefore a pair of subjects with at least one singleton is uncorrelated to each other. The covariance between singleton i ($i = n_{\mathrm{MZ}}+n_{\mathrm{DZ}}+1, \ldots, n$) and any other singleton or twin i' ($i' \neq i$) is

$$\begin{aligned}\mathbb{C}\mathrm{ovUN} &= \mathbb{C}\mathrm{ov}\begin{pmatrix}\mathrm{Y}_{i,0}\\ \mathrm{Y}_{i',j}\end{pmatrix}\\ &= \begin{pmatrix}A+C+E & 0\\ 0 & A+C+E\end{pmatrix},\end{aligned} \tag{3.4}$$

$j = 0, 1, \ldots, \frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})$. Finally, individuals in different twin pairs are mutually independent, so inter-twin covariance is also as given in Equation (3.4).

The covariance matrices for MZ twin pairs (3.2), DZ twin pairs (3.3) and unrelated subject pairs (3.4) can be further written in a linear fashion for a general algorithm

implementation:

$$\mathbb{C}\text{ovMZ} = A\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} + C\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} + E\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \tag{3.5}$$

$$\mathbb{C}\text{ovDZ} = A\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{pmatrix} + C\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} + E\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \tag{3.6}$$

$$\mathbb{C}\text{ovUN} = A\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + C\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + E\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}. \tag{3.7}$$

If we denote $\boldsymbol{\rho} = (A, C, E)^{\mathrm{T}}$ as a column vector consisting of these variance components, then we have a concise expression for the variance-covariance matrix $\mathbf{V}$:

$$\mathbf{V} = \sum_{l=1}^{3} \rho_l \mathbf{Q}_l,$$

where $\mathbf{Q}_l$ $(l = 1, 2, 3)$ is a block-diagonal matrix with blocks specified in Equations (3.5), (3.6) and (3.7), corresponding to the arrangement of MZ, DZ and singletons. For example, if $n_{\mathrm{MZ}} = 2,\ n_{\mathrm{DZ}} = 2,\ n_{\mathrm{S}} = 1$, then

$$\mathbf{Q}_1 = \begin{pmatrix}1 & 1 & 0 & 0 & 0\\ 1 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & \frac{1}{2} & 0\\ 0 & 0 & \frac{1}{2} & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix},\ \mathbf{Q}_2 = \begin{pmatrix}1 & 1 & 0 & 0 & 0\\ 1 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix},\ \mathbf{Q}_3 = \begin{pmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}.$$

3.2 Brief Review of the Existing Methods

3.2.1 Falconer's Method

The heritability method due to Falconer (Falconer and Mackay, 1996) is based on moment matching of intraclass correlation coefficients between MZ twins (r_{MZ}) and DZ twins (r_{DZ}):

$$\mathbb{E}\left[r_{\mathrm{MZ}}\right] = \frac{A + C}{A + C + E},$$
$$\mathbb{E}\left[r_{\mathrm{DZ}}\right] = \frac{\frac{1}{2}A + C}{A + C + E}.$$

Solving for narrow sense heritability (2.3), these equations give the Falconer's heritability estimator:

$$\widehat{h}^2_{\mathrm{F}} = 2 \times (r_{\mathrm{MZ}} - r_{\mathrm{DZ}}). \tag{3.8}$$

This method is the earliest and simplest way to estimate the heritability (Falconer and Mackay, 1996), but earlier work has shown that it performs worse than the likelihood-based methods (Nichols et al., 2009).

3.2.2 Bayesian Restricted Maximum Likelihood

Compared with Falconer's method, an implementation of restricted maximum likelihood (ReML) (Harville, 1977), a modified ikelihood-based estimation method by applying the maximum-likelihood principles to the residuals (see more details in Section 3.3.1), is embedded in the Statistical Parametric Mapping (SPM) software [1] in MATLAB. The SPM software package, including a suite of MATLAB functions and subroutines (with some externally compiled C routines), is designed for the brain imaging data analysis, and is freely available to the neuroimaging community. Previous studies showed that this Bayesian ReML approach in SPM is more accurate than Falconer's method in heritability estimation, producing the estimates with lower bias and smaller variance (Nichols et al., 2009).

SPM uses a non-standard ReML implementation, a Bayesian version of ReML, for the estimation of variance components (hyperparameters in the hierarchical model) with a Gaussian prior, where the log transformation can be further employed to enforce the non-negative constraints on variance components [2]. In Bayesian inference, the parameters are treated as random variables with a Gaussian distribution, and a linear hierarchical model for timeseries $\mathbf{Y}$ from all subjects can be constructed as

$$\mathbf{Y} = \mathbf{X}^{(1)}\boldsymbol{\theta}^{(1)} + \boldsymbol{\epsilon}^{(1)},$$

$$\boldsymbol{\theta}^{(1)} = \mathbf{X}^{(2)}\boldsymbol{\theta}^{(2)} + \boldsymbol{\epsilon}^{(2)},$$

where $\mathbf{X}^{(i)}$ is the specified design matrix at level i, $\boldsymbol{\theta}^{(i)}$ are parameters at level i, and the errors $\boldsymbol{\epsilon}^{(i)}$ at level i are distributed as $\mathbb{N}(\mathbf{0}, \mathbf{C}_{\boldsymbol{\epsilon}}^{(i)})$. The covariance matrix $\mathbf{C}_{\boldsymbol{\epsilon}}^{(i)}$ can be written as $\mathbf{C}_{\boldsymbol{\epsilon}}^{(i)} = \sum_j \rho_j^{(i)} \mathbf{Q}_j^{(i)}$, where the hyperparameters $\rho_j^{(i)}$ are variance

[1] `http://www.fil.ion.ucl.ac.uk/spm/`

[2] For reference, the SPM user-recommended configuration for the Gaussian prior distribution of log-hyperparameters is with hyperprior expectation $h\mathrm{E} = \log(\mathbb{V}\mathrm{ar}(\mathrm{Y})) - 1$ and hyperprior covariance $h\mathrm{C} = \exp(8)$, both of which were shown to be preferable to other settings in simulations in terms of estimation accuracy and model selection.

components at level i and $\mathbf{Q}_j^{(i)}$ is the corresponding basis set. This two-level model is identical to a non-hierarchical model:

$$\mathbf{Y} = \mathbf{X}^{(1)}\mathbf{X}^{(2)}\boldsymbol{\theta}^{(2)} + \mathbf{X}^{(1)}\boldsymbol{\epsilon}^{(2)} + \boldsymbol{\epsilon}^{(1)},$$

and can be re-written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}^{(1)},\ \mathbf{X} = [\mathbf{X}^{(1)}, \mathbf{X}^{(1)}\mathbf{X}^{(2)}],\ \boldsymbol{\theta} = [\boldsymbol{\epsilon}^{(2)}, \boldsymbol{\theta}^{(2)}]^{\mathrm{T}}.$$

Based on the conditional probability and Bayes' rule in probability theory, the posterior density of parameters is proportional to the product of the prior density and the likelihood function, and thus the conditional covariance $\mathbf{C}_{\boldsymbol{\theta}|\mathbf{Y}}$ and conditional expectation $\boldsymbol{\eta}_{\boldsymbol{\theta}|\mathbf{Y}}$ of the posterior density are derived as

$$\mathbf{C}_{\boldsymbol{\theta}|\mathbf{Y}} = (\overline{\mathbf{X}}^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\epsilon}}^{-1}\overline{\mathbf{X}})^{-1},$$

$$\boldsymbol{\eta}_{\boldsymbol{\theta}|\mathbf{Y}} = \mathbf{C}_{\boldsymbol{\theta}|\mathbf{Y}}(\overline{\mathbf{X}}^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\epsilon}}^{-1}\overline{\mathbf{Y}}),$$

where

$$\overline{\mathbf{X}} = [\mathbf{X}, \mathbf{I}]^{\mathrm{T}},\ \overline{\mathbf{Y}} = [\mathbf{Y}, \mathbb{E}[\boldsymbol{\theta}]]^{\mathrm{T}},\ \mathbf{C}_{\boldsymbol{\epsilon}} = \begin{pmatrix} \mathbf{C}_{\boldsymbol{\epsilon}}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbb{C}\mathrm{ov}(\boldsymbol{\theta}) \end{pmatrix},$$

and $\mathbb{C}\mathrm{ov}(\boldsymbol{\theta}^{(2)}) = \infty$ if $\boldsymbol{\theta}^{(2)}$ are treated as unknown or $\mathbb{C}\mathrm{ov}(\boldsymbol{\theta}^{(2)}) = \mathbf{0}$ if known.

The expectation-maximization (EM) algorithm is adopted to iteratively search for the maximum likelihood estimator for the parameters and hyperparameters jointly by maximising the ReML log-likelihood (the objective function). In general, the EM algorithm in this Bayesian framework can be partitioned into two steps: the E-step is to find the conditional expectation and covariance of the parameters with fixed hyperparameters; the M-step is to update the maximum likelihood estimate of the hyperparameters using Fisher scoring algorithm (see Section 3.3.2) by holding the parameters fixed. For linear models under the Gaussian assumptions, the EM scheme is equivalent to the classical ReML (Friston et al., 2002a,b).

3.2.3 Structural Equation Modeling

The freely accessible software "Mx" [3] and R package "OpenMx" [4] offer a specialized structural equation modeling (SEM) framework to allow flexible model definition

[3] `http://www.vcu.edu/mx/`

[4] `http://openmx.psyc.virginia.edu/`

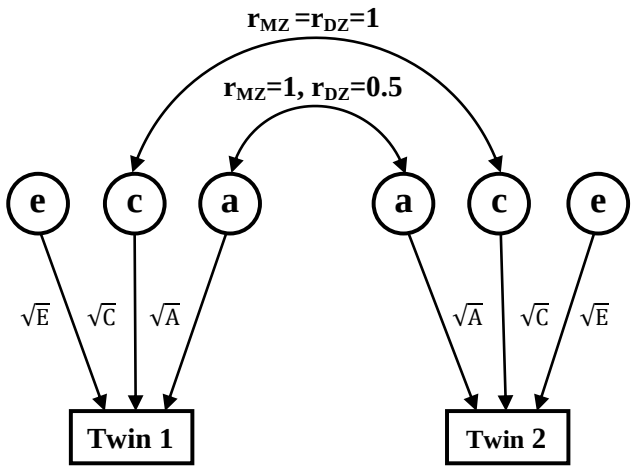


Figure 3.1: Path diagram for the univariate ACE twin model.

and parameter estimation for variance components, both of which are commonly used in analyzing genetic data for heritability inference. The SEM ACE model for univariate twin data can be commonly displayed as a path diagram, shown in Figure 3.1, where the influence caused by the latent variables a, c and e can be described by the path coefficients $\sqrt{A}$, $\sqrt{C}$ and $\sqrt{E}$ respectively (Rijsdijk and Sham, 2002). According to path tracing rules, the covariance matrices for MZ and DZ twin pairs are

$$\begin{pmatrix} A+C+E & A+C \\ A+C & A+C+E \end{pmatrix} \text{ and } \begin{pmatrix} A+C+E & \frac{1}{2}A+C \\ \frac{1}{2}A+C & A+C+E \end{pmatrix}$$

respectively, which have the same structure as matrices (3.2) and (3.3). The goodness of fit of this model is also measured using the maximum likelihood criterion (Rijsdijk and Sham, 2002). However, there exist some drawbacks of this SEM approach employed in Mx and OpenMx for the imaging data analysis. The goodness-of-fit likelihood ratio test statistic asymptotically follows a mixture of chi-square distributions (Self and Liang, 1987; Dominicus et al., 2006; Zhang and Lin, 2008), but Mx incorrectly uses a single standard chi-square distribution (Rijsdijk and Sham, 2002). Also the use of these softwares requires file conversion between different softwares (e.g., R and Matlab).

3.3 Frequentist Restricted Maximum Likelihood

The standard method for estimating variance components in the small sample problems is restricted maximum likelihood (ReML), which is also the default method

for variance components analysis in most statistical packages. ReML is simply applying maximum likelihood approach on least squares residuals. The variance component parameters are estimated using the Fisher scoring algorithm (Jennrich and Sampson, 1976; Harville, 1977). However, there are two issues with this numerical optimization problem, having to do with non-negativity and convergence, and thus this algorithm is further modified to both satisfy the non-negative constraint and improve the convergence of the algorithm.

The modified algorithm includes the logarithm parameterization of variance components and numerical optimization of the maximum likelihood function on the log domain of parameters to keep the non-negativity constraint. A novel reparameterization approach is adopted to alter the boundary of the parameter space seeking for a better convergence rate. We also consider applying the line search method to largely guarantee the increase of the likelihood for each iteration of the algorithm. We call this modified algorithm "Frequentist ReML".

3.3.1 Restricted Likelihood Maximum

Maximum likelihood estimation (MLE) is a well-known and widely used approach to estimating unknown parameters in statistical models, which has various extensions. When used in variance components inference, however, MLE produces biased estimates since MLE fails to take into account the loss in degrees of freedom resulting from estimating the nuisance fixed effect parameters (Harville, 1977), but its estimator is asymptotically unbiased. In contrast with MLE, ReML is an alternative form of MLE, which accounts for the loss of degrees of freedom and, in general, produces less biased variance component estimates than MLE. The ReML approach was first proposed and introduced by Patterson and Thompson (1971), and later reviewed and summarised by Harville (1977). Currently ReML has become the most commonly used means of variance component analysis.

The ReML log-likelihood, after removing the constant term, is expressed as

$$\ell(\boldsymbol{\rho}|\mathbf{Y}) = -\frac{1}{2}\Big[\log|\mathbf{V}| + \log|\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}| + (\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}})\Big], \quad (3.9)$$

where

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \arg\min_{\boldsymbol{\beta}}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{Y}$$

is the generalized least squares (GLS) estimator for $\boldsymbol{\beta}$ (Harville, 1977). Simplifying Equation (3.9) and denoting $\mathbf{R} = \mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}$ yield a simplified expression of ReML log-likelihood:

$$\ell(\boldsymbol{\rho}|\mathbf{Y}) = -\frac{1}{2}\big[\log|\mathbf{V}| + \log|\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}| + \mathbf{Y}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{R}\mathbf{Y}\big].$$

The ReML estimate is obtained by maximizing this ReML log-likelihood function. But the solution of this optimization problem is implicit and no analytical expressions can be achieved, so an iterative algorithm—Fisher scoring algorithm (FSA) is employed to numerically and iteratively approximate the optimal value.

3.3.2 Fisher Scoring Algorithm

Similar to Newton's method, FSA is used to successively approximate the solution of this maximization problem based on Taylor's theorem. By substituting the observed information matrix $\nabla^2\ell(\boldsymbol{\rho}|\mathbf{Y})$ with its expectation (i.e., Fisher information matrix), FSA is defined to numerically search for the ReML estimate with zero gradient of the ReML log-likelihood function. As the remainder of Taylor approximation tends to zero for increasing sample size, the derived ReML estimate approximates to a local maximum point.

Denote the gradient, observed information matrix and Fisher information matrix of the ReML log-likelihood as

$$\nabla\ell(\boldsymbol{\rho}|\mathbf{Y}) = \left(\frac{\partial\ell(\boldsymbol{\rho})}{\partial\rho_1}, \frac{\partial\ell(\boldsymbol{\rho})}{\partial\rho_2}, \frac{\partial\ell(\boldsymbol{\rho})}{\partial\rho_3}\right)^{\mathrm{T}},$$

$$\nabla^2\ell(\boldsymbol{\rho}|\mathbf{Y}) = \begin{pmatrix} -\frac{\partial^2\ell(\boldsymbol{\rho})}{\partial{\rho_1}^2} & -\frac{\partial^2\ell(\boldsymbol{\rho})}{\partial\rho_1\partial\rho_2} & -\frac{\partial^2\ell(\boldsymbol{\rho})}{\partial\rho_1\partial\rho_3} \\ -\frac{\partial^2\ell(\boldsymbol{\rho})}{\partial\rho_2\partial\rho_1} & -\frac{\partial^2\ell(\boldsymbol{\rho})}{\partial{\rho_2}^2} & -\frac{\partial^2\ell(\boldsymbol{\rho})}{\partial\rho_2\partial\rho_3} \\ -\frac{\partial^2\ell(\boldsymbol{\rho})}{\partial\rho_3\partial\rho_1} & -\frac{\partial^2\ell(\boldsymbol{\rho})}{\partial\rho_3\partial\rho_2} & -\frac{\partial^2\ell(\boldsymbol{\rho})}{\partial{\rho_3}^2} \end{pmatrix},$$

$$\mathcal{I}(\boldsymbol{\rho}|\mathbf{Y}) = \mathbb{E}[\nabla^2\ell(\boldsymbol{\rho}|\mathbf{Y})],$$

respectively. The first-order partial derivative of ReML log-likelihood with respect to ρ_l $(l = 1, 2, 3)$ is

$$\frac{\partial\ell(\boldsymbol{\rho}|\mathbf{Y})}{\partial\rho_l} = \frac{1}{2}\big[\mathbf{Y}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{R}\mathbf{Q}_l\mathbf{V}^{-1}\mathbf{R}\mathbf{Y} - \mathrm{tr}(\mathbf{R}\mathbf{Q}_l\mathbf{V}^{-1})\big],$$

the second-order derivative of the ReML log-likelihood with respect to ρ_l and $\rho_{l'}$ $(l, l' =$

$1, 2, 3)$ is

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\rho}|\mathbf{Y})}{\partial \rho_l \partial \rho_{l'}} &= \frac{1}{2}\Big[\mathrm{tr}(\mathbf{Q}_l\mathbf{V}^{-1}\mathbf{R}\mathbf{Q}_{l'}\mathbf{V}^{-1}\mathbf{R}) - \mathbf{Y}^{\mathrm{T}}(\mathbf{V}^{-1}\mathbf{R}\mathbf{Q}_{l'}\mathbf{V}^{-1}\mathbf{R}\mathbf{Q}_l \times \\ &\quad \mathbf{V}^{-1}\mathbf{R} + \mathbf{V}^{-1}\mathbf{R}\mathbf{Q}_l\mathbf{V}^{-1}\mathbf{R}\mathbf{Q}_{l'}\mathbf{V}^{-1}\mathbf{R})\mathbf{Y}\Big],\end{aligned}$$

and the $(l, l')^{\mathrm{th}}$ element of the Fisher information matrix is derived with the expression of

$$\{\mathcal{I}(\boldsymbol{\rho}|\mathbf{Y})\}_{ll'} = \frac{1}{2}\mathrm{tr}(\mathbf{Q}_l\mathbf{V}^{-1}\mathbf{R}\mathbf{Q}_{l'}\mathbf{V}^{-1}\mathbf{R}).$$

Thus we finally establish the recursion equation of FSA as

$$\boldsymbol{\rho}_{s+1} = \boldsymbol{\rho}_s + \mathcal{I}^{-1}(\boldsymbol{\rho}_s|\mathbf{Y})\nabla\ell(\boldsymbol{\rho}_s|\mathbf{Y}),$$

where $s \geq 0$ indexes iterations.

The initial value $\boldsymbol{\rho}_0$ (starting point) is pre-defined to be the product of sample variance and Falconer's estimates (see Section 3.2.1). The convergence of FSA is measured in terms of the absolute difference between the consecutive ReML log-likelihood values: $|\ell(\boldsymbol{\rho}_{s+1}|\mathbf{Y}) - \ell(\boldsymbol{\rho}_s|\mathbf{Y})|$, and the criterion of $|\ell(\boldsymbol{\rho}_{s+1}|\mathbf{Y}) - \ell(\boldsymbol{\rho}_s|\mathbf{Y})| < 10^{-4}$ is used to determine when to stop iterations. Under mild regularity conditions, the inverse matrix of Fisher information is the asymptotic variance of the ReML estimator for large samples.

3.3.3 Algorithm Modifications

Under certain regularity conditions, FSA converges in probability to the local maximum for convex log-likelihood (Osborne, 1992), however, the ReML log-likelihood can be non-convex and FSA may not converge (Mishchenko et al., 2008). In Figure 3.2, we illustrate a simulated example of the quadric-shaped ReML log-likelihood function with the true ACE parameter setting of $(A, C, E)^{\mathrm{T}} = (0, 0, 1)^{\mathrm{T}}$ and total variance $\sigma^2 = 1$, which shows the ReML log-likelihood values for valid parameter settings in the ACE parameter space of $\{(A, C, E)^{\mathrm{T}} | A, C, E \in [0, 1] \text{ and } A+C+E = 1\}$. Nevertheless, the estimate of the parameter vector can jump outside the parameter space and the ReML log-likelihood measures may not keep non-positive (i.e., the ReML likelihood values are not between 0 and 1), leading to the divergence or the inaccuracy of an algorithm. As a variant of Newton's method, FSA is also sensitive to the starting point chosen; FSA can fail to converge if the given initial value $\boldsymbol{\rho}_0$ is poor. For instance, when the starting point is near or at the boundary of

the parameter space, FSA is frequent to finally arrive at a point estimate outside the parameter space. Therefore, some possible modifications of the algorithm are proposed below, attempting to solve the convergence problems.

Line Search

The line search method can be applied and embedded in FSA in order to make sure an adequate increase in the objective function is achieved for each iteration. For the iteration s,

$$\boldsymbol{\rho}_{s+1} = \boldsymbol{\rho}_s + \alpha_s \mathbf{p}_s,$$

where $\mathbf{p}_s = \mathcal{I}^{-1}(\boldsymbol{\rho}_s|\mathbf{Y})\nabla\ell(\boldsymbol{\rho}_s|\mathbf{Y})$ denotes a decent search direction of the objective function, and the step length α_s is employed so as to achieve certain progress in the objective values, i.e., to ensure an increase in the ReML log-likelihood. The optimised choice of step length would be the exact step length obtained by solving the equation $\nabla\ell(\boldsymbol{\rho}_s + \alpha_s\mathbf{p}_s|\mathbf{Y}) = 0$ with fixed $\boldsymbol{\rho}_s$ and $\mathbf{p}_s$, but the full analytical expression of the maximum point is unachievable in this case. As the exact step length is impossible to derive, the inexact step length must be used.

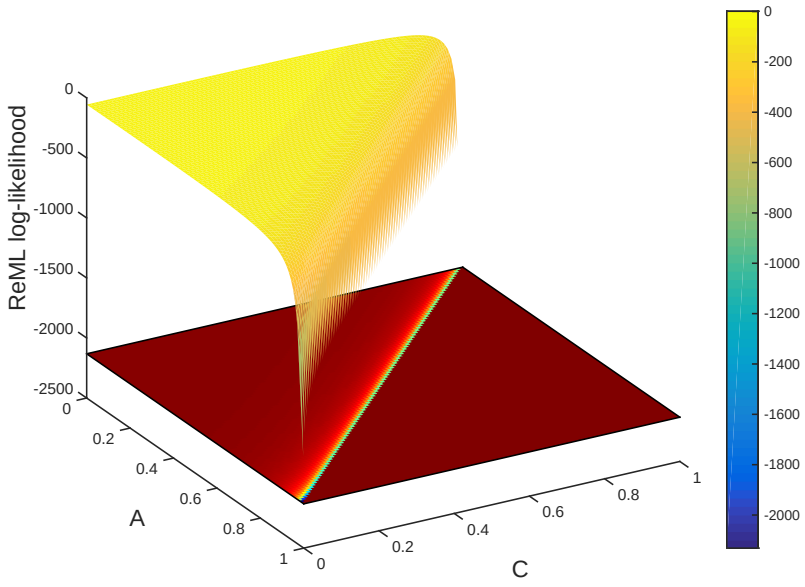
An acceptable inexact step length increases the ReML log-likelihood sufficiently, which can be determined by the Wolfe conditions (Wolfe, 1969, 1971). If $\alpha_s > 0$ satisfies the following two inequalities

$$\begin{cases} \ell(\boldsymbol{\rho}_s + \alpha_s\mathbf{p}_s|\mathbf{Y}) \geq \ell(\boldsymbol{\rho}_s|\mathbf{Y}) + c_1\alpha_s{\mathbf{p}_s}^{\mathrm{T}}\nabla\ell(\boldsymbol{\rho}_s|\mathbf{Y}), \\ \mathbf{p}_s^{\mathrm{T}}\nabla\ell(\boldsymbol{\rho}_s + \alpha_s\mathbf{p}_s|\mathbf{Y}) \leq c_2{\mathbf{p}_s}^{\mathrm{T}}\nabla\ell(\boldsymbol{\rho}_s|\mathbf{Y}), \end{cases}$$

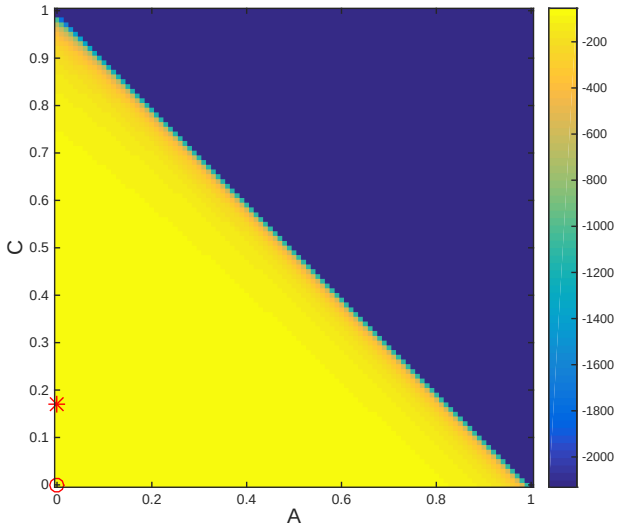
where $0 < c_1 < c_2 < 1$, then α_s is the optimal step size satisfying the Wolfe conditions. In practice, we use the recommended values of $c_1 = 10^{-4}$ and $c_2 = 0.9$ (Nocedal and Wright, 1999). For each iteration, the first inequality guarantees a sufficient increase of the objective function is obtained, and the second ensures that α_s sufficiently decreases the gradient of ReML log-likelihood—$\nabla\ell(\boldsymbol{\rho}_s|\mathbf{Y})$, so that it converges to zero (Wolfe, 1969, 1971). Our use of the backtracking rule for line search starts with an admissible choice of $\alpha_s = 1$, and repeats halving the step length with $\alpha_s = \frac{1}{2}\alpha_s$ until the Wolfe conditions are satisfied.

Logarithm Parameterization

The FSA algorithm can fail for some starting points and the reason is the constraints on the parameters, e.g., non-negativity. In order to guarantee that the variance



(a) ReML log-likelihood values on 3D surface.



(b) ReML log-likelihood values projected on 2D AC plane. The true parameter setting is marked in a red circle "○" and the parameters with the largest ReML log-likelihood are marked with a red star "✳".

Figure 3.2: An example of the ReML log-likelihood values for all valid parameters in the parameter space on (a) 3D surface and (b) 2D AC plane. Here the sample size is $n = 100$ with $n_{\text{MZ}} = 50$ MZ twins and $n_{\text{DZ}} = 50$ DZ twins. The true parameter setting is $(A, C, E)^{\text{T}} = (0, 0, 1)^{\text{T}}$ with the total variance of $\sigma^2 = 1$.

components are non-negative for each iteration, we further consider a logarithm parameterization of these variance components: $\boldsymbol{\rho} = \exp(\boldsymbol{\lambda})$, and the variance-covariance matrix $\mathbf{V}$ is written as

$$\mathbf{V} = \sum_{l=1}^{3} \exp(\lambda_l)\mathbf{Q}_l.$$

The numerical optimization of the ReML log-likelihood in terms of this log transformed parameter vector $\boldsymbol{\lambda}$ is implemented in the log domain. More specifically, the starting value of the variance components $\boldsymbol{\rho}$ is converted to provide the initial guess for $\boldsymbol{\lambda}$, FSA is applied to numerically search for the optimal values until iterating to convergence in the log scale, and the derived estimate of $\boldsymbol{\lambda}$ is exponentiated to derive the estimate of the parameter vector $\boldsymbol{\rho}$.

Reparametrization

While in real data applications we often found the C effect to be minor, it is not negligible since it improves the estimation accuracy and reduces the type I error rate for the A effect, i.e., the heritability. This small common environmental factor makes the parameter vector always close to the boundary of the parameter space, and thus, we consider a reparametrization of the variance components as a further attempt to improve convergence. The new reparametrization is defined as follows:

- Additional covariance of MZ twins: $\widetilde{A} = \frac{1}{2}A$,
- Overall minimum covariance: $\widetilde{C} = \frac{1}{2}A + C$,
- Unique environmental variance: $\widetilde{E} = E$.

In this way, the variance-covariance matrix $\mathbf{V}$ can be written as

$$\mathbf{V} = \sum_{l=1}^{3} \widetilde{\rho}_l \widetilde{\mathbf{Q}}_l,$$

where $\widetilde{\boldsymbol{\rho}} = (\widetilde{A}, \widetilde{C}, \widetilde{E})^{\mathrm{T}}$, and $\widetilde{\mathbf{Q}}_l$ $(l = 1, 2, 3)$ can be analogously constructed like $\mathbf{Q}_l$ in the ACE parameterization (see Section 3.1). The above-mentioned approaches for the ACE parameterization can also be applied to make inference on this $\widetilde{\mathrm{A}}\widetilde{\mathrm{C}}\widetilde{\mathrm{E}}$ reparametrization.

3.4 Linear Regression with Squared Differences

In the 1980s, a simple linear regression method for variance components estimation using squared differences (SD's) of each subject pair was proposed by Grimes and Harvey (1980), which is new for neuroimaging studies. For a sample of n subjects, there are totally n^2-n possible SD's, $\frac{1}{2}(n^2-n)$ unique and distinct. This SD analysis is based on the derivation that the variance of the difference between two correlated random variables depends on the variance-covariance parameters in a linear fashion, and the expectation of the SD's can be expressed in terms of variance components A, C and E (Grimes and Harvey, 1980; Lindquist et al., 2012). However, in Grimes and Harvey's paper, the non-negative restriction for variance components and heritability can not be satisfied with the ordinary least squares, and the derived negative heritability estimates were either eliminated (in their simulation study) or accepted as the impractical estimates during real data analysis, which results in this linear regression method being an inaccurate method for heritability estimation.

To deal with the non-negativity problem, Lawson and Hanson proposed a now ubiquitous solution called non-negative least squares (NNLS) (Lawson and Hanson, 1987). The foundation of this algorithm is the Karush-Kuhn-Tucker (KKT) conditions (Karush, 1939; Kuhn and Tucker, 1951), which were first proposed for the complex non-linear programming problems (Lawson and Hanson, 1987). In our linear case, the KKT conditions can be further simplified to accelerate the computation. Although other methods had been proposed to solve this non-negativity problem for large and sparse matrix settings, Luo and Duraiswami (2011) suggested that NNLS still maintained its superiority when small or moderate dense matrices were handled.

While Grimes and Harvey's method specifies a linear regression model with the use of $\frac{1}{2}(n^2 - n)$ different observations of SD's, our modification of this method simplifies the computation so that only $\frac{1}{2}(n_{\mathrm{MZ}} + n_{\mathrm{DZ}})$ observations are utilized in computing SD's. Thus, the integration of constructing linear regression model with SD's and estimating parameters using NNLS with computational modification yields a novel and fast non-negative least squares regression approach for unknown variance components estimation, entitled "LR-SD".

3.4.1 Linear Regression Model

The means of linear regression model construction with SD's varies depending on whether the between-subject covariates are included or not in the GLM model (3.1).

For the simplest case, where there is an intercept only, we call the model *simple linear regression*. While in a bit more complicated case, we call the model *multiple linear regression* when some covariates are considered.

Simple Linear Regression

Consider the case where our original GLM model (3.1) is a trivial simple linear regression model with an intercept coefficient only:

$$\mathbf{Y} = \mathbf{1}\beta_0 + \boldsymbol{\epsilon}, \tag{3.10}$$

where $\mathbf{1}$ is an all-ones vector and β_0 is the population mean. By the extension of the covariance matrices (3.2), (3.3) and (3.4) and the basic properties of variance operator, we get

$$\begin{aligned} \mathbb{E}\left[(\mathrm{Y}_{2j-1,j} - \mathrm{Y}_{2j,j})^2\right] &= \mathbb{V}\mathrm{ar}(\epsilon_{2j-1,j} - \epsilon_{2j,j}) \\ &= 2E \end{aligned} \tag{3.11}$$

for MZ twin pair j ($j = 1, \ldots, \frac{1}{2}n_{\mathrm{MZ}}$),

$$\begin{aligned} \mathbb{E}\left[(\mathrm{Y}_{2j-1,j} - \mathrm{Y}_{2j,j})^2\right] &= \mathbb{V}\mathrm{ar}(\epsilon_{2j-1,j} - \epsilon_{2j,j}) \\ &= A + 2E \end{aligned} \tag{3.12}$$

for DZ twin pair j ($j = \frac{1}{2}n_{\mathrm{MZ}}+1, \ldots, \frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})$), and for the remaining unrelated subject pair of $\mathrm{Y}_{i,j}$ and $\mathrm{Y}_{i',j'}$,

$$\begin{aligned} \mathbb{E}\left[(\mathrm{Y}_{i,j} - \mathrm{Y}_{i',j'})^2\right] &= \mathbb{V}\mathrm{ar}(\epsilon_{i,j} - \epsilon_{i',j'}) \\ &= 2A + 2C + 2E. \end{aligned} \tag{3.13}$$

These relationships (3.11), (3.12) and (3.13) describe the expected values for all these $\frac{1}{2}(n^2 - n)$ SD's and specify the mean structure of a linear regression model

with SD's as the observations:

$$\mathbb{E}\begin{pmatrix} (\mathrm{Y}_{1,1}-\mathrm{Y}_{2,1})^2 \\ \vdots \\ (\mathrm{Y}_{n_{\mathrm{MZ}}-1,\frac{1}{2}n_{\mathrm{MZ}}}-\mathrm{Y}_{n_{\mathrm{MZ}},\frac{1}{2}n_{\mathrm{MZ}}})^2 \\ (\mathrm{Y}_{n_{\mathrm{MZ}}+1,\frac{1}{2}n_{\mathrm{MZ}}+1}-\mathrm{Y}_{n_{\mathrm{MZ}}+2,\frac{1}{2}n_{\mathrm{MZ}}+1})^2 \\ \vdots \\ (\mathrm{Y}_{n_{\mathrm{MZ}}+n_{\mathrm{DZ}}-1,\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})}-\mathrm{Y}_{n_{\mathrm{MZ}}+n_{\mathrm{DZ}},\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})})^2 \\ (\mathrm{Y}_{1,1}-\mathrm{Y}_{3,2})^2 \\ \vdots \\ (\mathrm{Y}_{n-1,0}-\mathrm{Y}_{n,0})^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 2 \\ & \vdots & \\ 0 & 0 & 2 \\ 1 & 0 & 2 \\ & \vdots & \\ 1 & 0 & 2 \\ 2 & 2 & 2 \\ & \vdots & \\ 2 & 2 & 2 \end{pmatrix} \begin{pmatrix} A \\ C \\ E \end{pmatrix},$$

where the first $\frac{1}{2}n_{\mathrm{MZ}}$ SD's are derived from MZ twin pairs and based on Equation (3.11), the following $\frac{1}{2}n_{\mathrm{DZ}}$ elements are from DZ twin pairs and based on Equation (3.12), and the remaining are from the unrelated subject pairs and based on Equation (3.13). For simplicity, this linear regression model with SD's can be denoted as

$$\mathbb{E}\left[\mathbf{D}\right] = \mathbf{Z}\boldsymbol{\rho}, \tag{3.14}$$

where $\mathbf{D}$ is a $\frac{1}{2}(n^2-n)$-vector of $\frac{1}{2}(n^2-n)$ observations (SD's), $\mathbf{Z}$ is a pre-defined $\frac{1}{2}(n^2-n)\times 3$ design matrix as shown above, and $\boldsymbol{\rho}$ is the parameter vector of variance components.

Multiple Linear Regression

Now suppose that the GLM model (3.1) is a multiple regression model containing an regression intercept and multiple covariates, expressed as

$$\mathbf{Y} = \mathbf{1}\beta_0 + \mathbf{X}_1\beta_1 + \cdots + \mathbf{X}_{p-1}\beta_{p-1} + \boldsymbol{\epsilon}, \tag{3.15}$$

where the n-vectors of $\mathbf{X}_1, \ldots, \mathbf{X}_{p-1}$ are regressors, each associated with one of the $p-1$ covariates, and $\beta_1, \ldots, \beta_{p-1}$ are the corresponding regression coefficients. Even though the parameter vector $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})^{\mathrm{T}}$ is not of interest and we treat its elements as nuisance parameters in variance component analysis, it must be accounted for during the analysis. If we estimate these parameters by ordinary least squares (OLS), the resulting OLS estimator for $\boldsymbol{\beta}$ is expressed as

$$\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y},$$

where $\mathbf{X} = (\mathbf{1}, \mathbf{X}_1, \ldots, \mathbf{X}_{p-1})$ is the complete design matrix, and thus the OLS residual is

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\text{OLS}} = \left[\mathbf{I} - \mathbf{X}(\mathbf{X}^{\text{T}}\mathbf{X})^{-1}\mathbf{X}^{\text{T}}\right]\mathbf{Y}. \tag{3.16}$$

Denote a symmetric and idempotent matrix as $\mathbf{R} = \mathbf{I} - \mathbf{X}(\mathbf{X}^{\text{T}}\mathbf{X})^{-1}\mathbf{X}^{\text{T}}$. The OLS residual $\mathbf{e} = \mathbf{RY} = \mathbf{R}\boldsymbol{\epsilon}$ follows a normal distribution with mean $\mathbb{E}[\mathbf{e}] = \mathbf{0}$ and variance $\mathbb{C}\text{ov}(\mathbf{e}) = \mathbf{RVR}$, i.e., $\mathbf{e} \sim \mathbb{N}(\mathbf{0}, \mathbf{RVR})$, where the projection matrix $\mathbf{R}$ projects the unobservable error vector $\boldsymbol{\epsilon}$ to its estimate $\mathbf{e}$ that is orthogonal to the column space spanned by the columns of design matrix $\mathbf{X}$.

Covariates contribute to the between-subject variation, however, these "contributions" may confound the heritability estimation and bias the estimator. Hence it is important to model all known covariate-related effects in $\mathbf{X}$, and subsequently we base all heritability analyses on $\mathbf{e}$, calculated with Equation (3.16). For simplicity of notation going forward, we write $\mathbf{Y}$ instead of $\mathbf{e}$ for the covariate-adjusted data. This embodies an assumption that we can estimate covariate effects $\boldsymbol{\beta}$ with high precision, and that $n \gg p$ so that $\mathbf{V}$ approximates the actual covariance matrix $\mathbf{RVR}$ of the adjusted data; we neglect any correlation induced by removing covariates and mean centering, and omit the lost information due to the loss of degrees of freedom caused by OLS estimation. Thus $\mathbf{Y}$ has mean $\mathbf{0}$ during the subsequent linear regression model construction with SD's.

Hence the expectations for different subject pairs are similarly derived as

$$\begin{aligned}\mathbb{E}\left[(\text{Y}_{2j-1,j} - \text{Y}_{2j,j})^2\right] &= \mathbb{V}\text{ar}(\text{Y}_{2j-1,j} - \text{Y}_{2j,j}) \\ &\approx 2E\end{aligned}$$

for MZ twin pair j ($j = 1, \ldots, \frac{1}{2}n_{\text{MZ}}$);

$$\begin{aligned}\mathbb{E}\left[(\text{Y}_{2j-1,j} - \text{Y}_{2j,j})^2\right] &= \mathbb{V}\text{ar}(\text{Y}_{2j-1,j} - \text{Y}_{2j,j}) \\ &\approx A + 2E\end{aligned}$$

for DZ twin pair j ($j = \frac{1}{2}n_{\text{MZ}}+1, \ldots, \frac{1}{2}(n_{\text{MZ}}+n_{\text{DZ}})$); and for the remaining unrelated subject pair of $\text{Y}_{i,j}$ and $\text{Y}_{i',j'}$,

$$\begin{aligned}\mathbb{E}\left[(\text{Y}_{i,j} - \text{Y}_{i',j'})^2\right] &= \mathbb{V}\text{ar}(\text{Y}_{i,j} - \text{Y}_{i',j'}) \\ &\approx 2A + 2C + 2E.\end{aligned}$$

The deriving linear regression model with SD's in this case can be analogously denoted as

$$\mathbb{E}[\mathbf{D}] \approx \mathbf{Z}\boldsymbol{\rho}. \tag{3.17}$$

3.4.2 Non-negative Least Squares

Our LR-SD method proceeds by applying NNLS algorithm to the linear regression model with SD's (3.14) or (3.17) for the unknown variance components estimation; precisely, we seek

$$\min_{\boldsymbol{\rho}} f(\boldsymbol{\rho}) \quad \text{s.t.} \quad \boldsymbol{\rho} \geq \mathbf{0}, \tag{3.18}$$

where $f(\boldsymbol{\rho}) = \frac{1}{2}\|\mathbf{Z}\boldsymbol{\rho} - \mathbf{D}\|^2$ is the objective function to be minimized. Karush-Kuhn-Tucker (KKT) conditions provide the necessary conditions for this optimization problem: If $\boldsymbol{\rho}^*$ is the local minimizer of $f(\boldsymbol{\rho})$ satisfying the inequality constraint: $\boldsymbol{\rho} \geq 0$, then the following conditions hold:

$$\nabla f(\boldsymbol{\rho}^*)^{\mathrm{T}}\boldsymbol{\rho}^* = 0, \quad \nabla f(\boldsymbol{\rho}^*) \geq \mathbf{0}, \quad \boldsymbol{\rho}^* \geq \mathbf{0},$$

where the gradient vector is $\nabla f(\boldsymbol{\rho}) = \mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\rho} - \mathbf{D})$ (Karush, 1939; Kuhn and Tucker, 1951). As $\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\rho} - \mathbf{D}) = \mathbf{0}$ corresponds to the least squares normal equation, for any $\mathbf{Z}$, $\mathbf{D}$ and $\boldsymbol{\rho}$ found by least squares, the first two conditions are trivially satisfied.

Algorithm Simplification

The NNLS algorithm can be further modified and simplified for computation in our case since there are only 3 parameters A, C and E in total. Inclusion and exclusion of these parameters for minimization of the objective function lead to totally $2^3 - 1 = 7$ parameter vectors for the non-negativity selection. As the unique environmental factor E always exists due to the unavoidable measurement error, only 4 possible models E, AE, CE and ACE are taken into consideration finally, with parameter vectors denoted as $\boldsymbol{\rho}_{\mathrm{E}} = (0, 0, E)^{\mathrm{T}}$, $\boldsymbol{\rho}_{\mathrm{AE}} = (A, 0, E)^{\mathrm{T}}$, $\boldsymbol{\rho}_{\mathrm{CE}} = (0, C, E)^{\mathrm{T}}$ and $\boldsymbol{\rho}_{\mathrm{ACE}} = (A, C, E)^{\mathrm{T}}$ respectively.

Since the space of possible models is so small, we can enumerate and evaluate all these 4 models. If all parameter estimates of the full ACE model are non-negative, then $\widehat{\boldsymbol{\rho}}_{\mathrm{ACE}}$ is selected as the estimate $\widehat{\boldsymbol{\rho}}$ solving the above-mentioned NNLS problem (3.18); otherwise, the 3 remaining models corresponding to the parameter vectors $\boldsymbol{\rho}_{\mathrm{E}}$, $\boldsymbol{\rho}_{\mathrm{AE}}$, and $\boldsymbol{\rho}_{\mathrm{CE}}$ are compared. Among the optional models out of 3 with valid estimates (i.e., all non-negative), we select the NNLS estimate as the one derived

from the more complicated model with more dimensions considered, which has more variance components taken into consideration and involved in parameter estimation, and considers a larger parameter space. If both of the AE and CE models have valid estimates, we compute their OLS residual sum of squares (RSS) and choose the one with the smallest RSS value. Alternatively, we could compute the ReML log-likelihood of the original GLM model (3.1) and select the valid estimate that has the largest ReML log-likelihood value. We call the former method "LR-SD", and the latter "LR-SD ReML". But this simplification does not address the issue that for large n, the computation and storage of the length $\frac{1}{2}(n^2 - n)$ data and design matrix are unwieldy.

Computational Modification

In NNLS algorithm, we make use of the regular structure of $\mathbf{D}$ (the SD vector) to directly compute the least squares estimates. Specifically, observe that

$$\mathbf{Z}^{\mathrm{T}}\mathbf{D} = \begin{pmatrix} \sum_{l=\frac{1}{2}n_{\mathrm{MZ}}+1}^{\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})} \mathrm{D}_l + 2\sum_{l=\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})+1}^{\frac{1}{2}(n^2-n)} \mathrm{D}_l \\ 2\sum_{l=\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})+1}^{\frac{1}{2}(n^2-n)} \mathrm{D}_l \\ 2\sum_{l=1}^{\frac{1}{2}(n^2-n)} \mathrm{D}_l \end{pmatrix},$$

where D_l denotes the l^{th} element of $\mathbf{D}$. We can avoid computing all possible differences, needed for the third element, if we note that in *simple linear regression*,

$$\begin{aligned} \sum_{l=1}^{\frac{1}{2}(n^2-n)} \mathrm{D}_l &= \frac{1}{2}\sum_{l=1}^{n}\sum_{l'=1}^{n}(\mathrm{Y}_l - \mathrm{Y}_{l'})^2 \\ &= \frac{1}{2}\sum_{l=1}^{n}\sum_{l'=1}^{n}[(\mathrm{Y}_l - \overline{\mathrm{Y}}) + (\overline{\mathrm{Y}} - \mathrm{Y}_{l'})]^2 \\ &= n\sum_{l=1}^{n}(\mathrm{Y}_l - \overline{\mathrm{Y}})^2 \\ &= (n^2 - n) \times \mathrm{s}^2(\mathbf{Y}), \end{aligned}$$

where the sample variance of the n observations,

$$\mathrm{s}^2(\mathbf{Y}) = \frac{1}{n-1}(\mathrm{Y}_l - \overline{\mathrm{Y}})^2,$$

is an estimator for the phenotypic variance σ^2 with $n-1$ degrees of freedom. While in *multiple linear regression*, the sum can be similarly approximated with the OLS RSS value derived from the GLM model (3.1), i.e.,

$$\sum_{l=1}^{\frac{1}{2}(n^2-n)} \mathrm{D}_l \approx (n^2-n) \times \widehat{\sigma}^2,$$

where

$$\widehat{\sigma}^2 = \frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n-p}$$

is the OLS estimator for the variance σ^2 with $n-p$ degrees of freedom. Thus, we have modified and derived a simpler computational algorithm so that only $\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})$ rather than $\frac{1}{2}(n^2-n)$ observations of SD's are used.

3.4.3 Likelihood Ratio Test

Tests on parameter estimates are performed as usual, with a likelihood ratio test (LRT) comparing the fitted model (alternative model) with the constraint $\mathrm{H}_1 : A \geq 0$ to the null model with the hypothesis $\mathrm{H}_0 : A = 0$ (i.e., the model ACE is compared to CE, and AE is compared to E). The LRT statistic, termed as "T", is defined as

$$\mathrm{T} = -2 \times [\ell(\widehat{\boldsymbol{\rho}}_0|\mathbf{Y}) - \ell(\widehat{\boldsymbol{\rho}}_1|\mathbf{Y})],$$

where $\widehat{\boldsymbol{\rho}}_0$ and $\widehat{\boldsymbol{\rho}}_1$ are parameter estimates derived from the null model and the alternative model respectively, and proven to asymptotically follow a chi-squared distribution with 1 degree of freedom, i.e., χ^2_1 (Wilks, 1938). However the variance parameter A lies on the boundary of the parameter space of $\boldsymbol{\rho} = (A, C, E)^{\mathrm{T}}$ under the null hypothesis $\mathrm{H}_0 : h^2 = 0$ or, equivalently, $\mathrm{H}_0 : A = 0$, and thus the asymptotic sampling distribution of this LRT statistic, assuming H_0 is true, is a mixture of chi-squared distributions $\left(\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1\right)$ instead of a standard chi-square distribution (χ^2_1) (Self and Liang, 1987; Dominicus et al., 2006; Zhang and Lin, 2008).

Given the asymptotic null distribution of the LRT statistic, the theoretical p-value can be easily calculated. Obtaining a p-value less than a given significance level α, which is typically a small number (e.g., $\alpha = 0.05$), suggests that there is a significant evidence against the null hypothesis and the null hypothesis should be rejected at level α. The p-value, denoted as p, also indicates there is a $100p\%$ risk of incorrectly rejecting a true null hypothesis. Aside from the asymptotic theoretical p-value, the permutation-based p-value is exact, based on the empirical distribution of the LRT

statistic.

3.4.4 Relation to U-statistic

The use of U-statistics is an effective way of obtaining unbiased estimators, and even the minimum-variance unbiased estimators. Assume the existence of a sequence of n i.i.d. random variables (or random vectors) $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$ from an unknown population. For a positive integer m ($m \leq n$), the unknown parameter θ that is required to be estimated is of the form

$$\theta = \mathbb{E}\left[h(\mathrm{Y}_{i_1}, \ldots, \mathrm{Y}_{i_m})\right]$$

with a certain function $h(\cdot)$, which is assumed to be symmetric of its m arguments and gives rise to the U-statistic associated with the estimable parameter θ. The U-statistic corresponding to this symmetric function $h(\cdot)$ of order m is defined as

$$\mathrm{U}_n = \binom{n}{m}^{-1} \sum \cdots \sum_{(i_1,\ldots,i_m)\in\Omega_m^n} h(\mathrm{Y}_{i_1}, \ldots, \mathrm{Y}_{i_m}),$$

where n is the sample size, $h(\cdot)$ is called the kernel function, and

$$\Omega_m^n = \{(i_1, \ldots, i_m) | 1 \leq i_1 < \cdots < i_m \leq n\}$$

is a set of all possible combinations of m different elements from $\{1, \ldots, n\}$.

For each pair of twins regardless of MZ or DZ type with paired data $(\mathrm{Y}_{2j-1,j}, \mathrm{Y}_{2j,j})^{\mathrm{T}}$ for twin pair j, let $m = 1$ and the kernel function be

$$h\left(\begin{pmatrix} \mathrm{Y}_{2j-1,j} \\ \mathrm{Y}_{2j,j} \end{pmatrix}\right) = \frac{1}{2}(\mathrm{Y}_{2j-1,j} - \mathrm{Y}_{2j,j})^2.$$

Based on Equations (3.11) and (3.12) for MZ and DZ twins, we obtain

$$\mathbb{E}\left[h\left(\begin{pmatrix} \mathrm{Y}_{2j-1,j} \\ \mathrm{Y}_{2j,j} \end{pmatrix}\right)\right] = \begin{cases} E, & \text{for MZ twins,} \\ \frac{1}{2}A + E, & \text{for DZ twins.} \end{cases}$$

As distinct twin pairs are mutually independent, the U-statistic associated with E using the data from MZ twins is

$$\mathrm{U}_{n_{\mathrm{MZ}}} = \frac{1}{n_{\mathrm{MZ}}} \sum_{j=1}^{\frac{1}{2}n_{\mathrm{MZ}}} (\mathrm{Y}_{2j-1,j} - \mathrm{Y}_{2j,j})^2, \tag{3.19}$$

and the U-statistic corresponding to $\frac{1}{2}A + E$ with the data from DZ twins is

$$\mathrm{U}_{n_{\mathrm{DZ}}} = \frac{1}{n_{\mathrm{DZ}}} \sum_{j=\frac{1}{2}n_{\mathrm{MZ}}+1}^{\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})} (\mathrm{Y}_{2j-1,j} - \mathrm{Y}_{2j,j})^2. \tag{3.20}$$

With the use of Equations (3.19) and (3.20) as the unbiased estimators for E and $\frac{1}{2}A + E$ respectively, we derive the unbiased estimators for A and E to be

$$\widehat{A} = \frac{2}{n_{\mathrm{DZ}}} \sum_{j=\frac{1}{2}n_{\mathrm{MZ}}+1}^{\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})} (\mathrm{Y}_{2j-1,j} - \mathrm{Y}_{2j,j})^2 - \frac{2}{n_{\mathrm{MZ}}} \sum_{j=1}^{\frac{1}{2}n_{\mathrm{MZ}}} (\mathrm{Y}_{2j-1,j} - \mathrm{Y}_{2j,j})^2,$$

$$\widehat{E} = \frac{1}{n_{\mathrm{MZ}}} \sum_{j=1}^{\frac{1}{2}n_{\mathrm{MZ}}} (\mathrm{Y}_{2j-1,j} - \mathrm{Y}_{2j,j})^2,$$

which are identical with those estimators for A and E derived using the LR-SD method.

3.5 Simulation Studies

In this section, the 1D simulation analysis is conducted to compare our newly proposed voxel-wise univariate heritability estimation methods with those frequently used methods in terms of prediction accuracy, validity, sensitivity, and the overall computation time for different variance component settings. The ROC-based simulation studies generate 2D imaging data for power evaluation and comparison between the voxel- and cluster-wise heritability inference approaches with various simulation settings.

3.5.1 Simulation Evaluations

In order to evaluate our proposed heritability estimation methods—Frequentist ReML, LR-SD and LR-SD ReML, the Monte Carlo simulation is undertaken by fitting the GLM model (3.1), or linear regression model with SD's (3.14) or (3.17) to the 1D simulated datasets generated with different $(A, C, E)^{\mathrm{T}}$ parameter settings.

Simulation Setting

The parameter settings shown in Table 3.1 are motivated as follows. If we create a 3D Cartesian coordinate system with x, y and z axes representing the possible values for A, C and E, then the parameter space is formed and can be visualized

Table 3.1: 15 parameter settings of variance components $(A, C, E)^{\mathrm{T}}$.

	A	C	E
Complete Null	0	0	1
Only Common Environmental	0	1/6	5/6
Component	0	1/3	2/3
($A = 0$ & $C > 0$)	0	1/2	1/2
	0	2/3	1/3
Only Genetic Component	1/6	0	5/6
($A > 0$ & $C = 0$)	1/3	0	2/3
	1/2	0	1/2
	2/3	0	1/3
Both Genetic and Common	1/6	1/6	2/3
Environmental Components	1/3	1/6	1/2
($A > 0$ & $C > 0$)	1/6	1/3	1/2
	1/2	1/6	1/3
	1/3	1/3	1/3
	1/6	1/2	1/3

as an equilateral triangle. Within this equilateral triangular parameter space, a 2D Barycentric coordinate system can be set up by assigning a vertex of the triangle to be the origin, as shown in Figure 3.3. In reality, $E \gg \max(A, C)$ always holds, so we can evenly choose 15 possible sets of $(A, C, E)^{\mathrm{T}}$ from the upper part of this parameter space (the equilateral triangle) satisfying $E \geq \frac{1}{3}$ for simulation analysis, shown in Figure 3.3. Define the total variance as $\sigma^2 = A + C + E > 0$. Since we are more concerned about heritability, the relative portion of A over the sum of A, C and E. Without loss of generality, A, C and E can be scaled such that $\sigma^2 = 1$; that is, the value of A is exactly h^2. However, we still use

$$\widehat{h}^2 = \frac{\widehat{A}}{\widehat{A} + \widehat{C} + \widehat{E}}$$

during computation to account for the genetic random variation in total variance.

There are 3 samples considered with the size of $n = 20, 60, 100$, each comprised of twins only, where the number of MZ twins is assumed to be identical with that

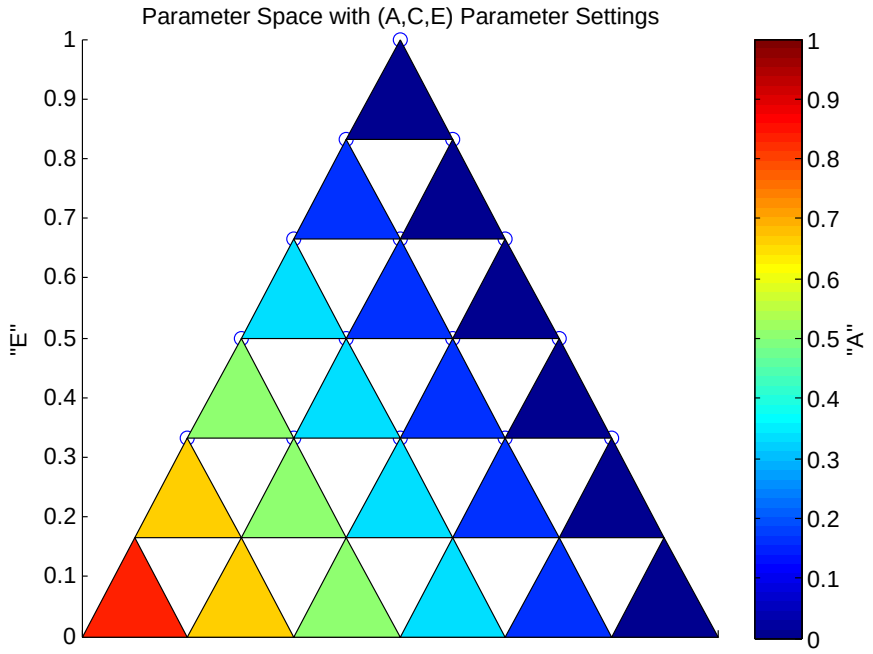


Figure 3.3: Parameter space with various circled $(A, C, E)^{\mathrm{T}}$ parameter settings. The large equilateral triangle depicts the ACE parameter space with y axis representing the values of E and circled points representing the selected $(A, C, E)^{\mathrm{T}}$ parameter settings shown in Table 3.1. The colored small triangles indicate the values of A for the top vertices.

of DZ twins. For instance, the sample of 20 subjects is comprised of 5 MZ twin pairs (10 subjects) and 5 DZ twin pairs (10 subjects). In summary, 3 different sample sizes, along with 15 parameter settings of $(A, C, E)^{\mathrm{T}}$, lead to the 45 1D simulation settings. For each simulation setting, we considered both cases of the GLM model (3.1)—*simple linear regression* (3.10) and *multiple linear regression* (3.15). For *simple linear regression*, the design matrix is the all-ones vector, i.e., $\mathbf{X} = (1, \ldots, 1)^{\mathrm{T}}$. For *multiple linear regression*, the design matrix $\mathbf{X}$ is generated to have two columns, where the first column is an all-ones vector and the second is a standard uniformly distributed random vector approximating the covariates. Totally $n_{\mathrm{Rlz}} = 10000$ simulations were executed.

Comparison Results

We make a comparison between our proposed methods of Frequentist ReML, LR-SD and LR-SD ReML and those existing methods including Falconer's method,

Bayesian ReML used in SPM and SEM used in OpenMx. The statistical measures including mean squared error (MSE), false positive rate (FPR) and statistical power of LRT, and overall running time of $n_{\text{Rlz}} = 10000$ simulations for these methods are used for comparative purposes. Since both cases of *simple linear regression* and *multiple linear regression* reflect strongly similar simulation results, we will only illustrate the results obtained from *simple linear regression* for simplicity.

Accuracy and Precision The MSE comparison of these 6 heritability estimation methods is shown in Figure 3.4, which exhibits that the two linear regression methods—LR-SD and LR-SD ReML, have nearly identical MSE measures in all simulation settings, implying that they provide heritability estimates with resembling accuracy. The method of SEM in OpenMx also shows strong similarity of the MSE performance compared with the linear regression methods. Nevertheless, both Frequentist ReML and Bayesian ReML in SPM show larger MSE values for the first 5 null settings with $A = 0$ in the smallest sample case of size $20 = 10 + 10$, especially for Frequentist ReML. As expected, the MSE in heritability estimator using Falconer's method always exceeds that using the other 5 methods, which is as reported in previous results (Nichols et al., 2009).

Statistical Sensitivity Figure 3.5 shows the statistical power comparison of LRT using these 6 methods at the nominal significance level $\alpha = 0.05$ under the null hypothesis of no heritability. The validity of all these methods is verified except Frequentist ReML. Although Frequentist ReML has the largest statistical power overall, its FPR is above the upper confidence bound, which means Frequentist ReML is not applicable due to its high risk of wrongly rejecting the true null hypothesis with LRT. It should also be noted that the estimation inaccuracy and high FPR of Frequentist ReML may be due to the saddle point problem and need to be further investigated, where a saddle point is a stationary point but not a local extremum of a function. For the remaining 5 methods, they are all within or below the confidence limits. Even if these 5 estimation methods similarly return low power when small samples are dealt with, increasing the sample size results in a notably enlarged statistical power. For nearly all simulation settings considered, LR-SD, LR-SD ReML and SEM in OpenMx perform fairly similar, and their statistical power exceeds that of Falconer's method and Bayesian ReML in SPM, particularly for the parameter settings with low values of E. The estimated FPR is employed to assess the validity for all these methods in Figure 3.6, where laying in or below the 95% binomial proportion confidence interval (plotted as the region between the two red

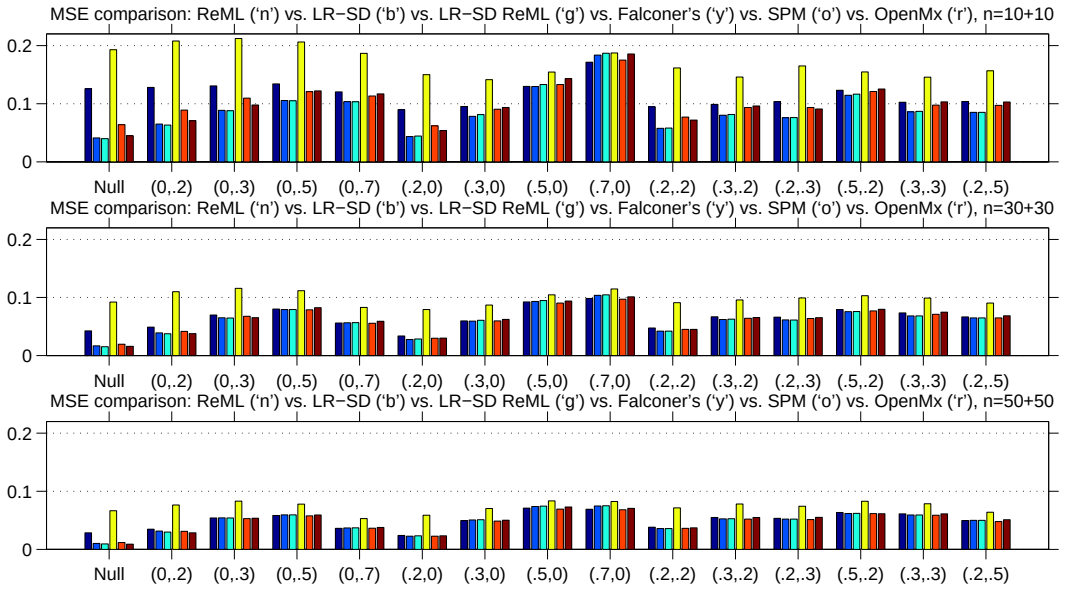


Figure 3.4: The MSE comparison: Frequentist ReML ('n': navy), LR-SD ('b': blue), LR-SD ReML ('g': green), Falconer's method ('y': yellow), Bayesian ReML used in SPM ('o': orange), and SEM used in OpenMx ('r': red). Comma ordered pairs on x-axis correspond to the rounded parameter values of A and C, i.e., (A, C); see Table 3.1 and Figure 3.3 for exact parameter settings used.

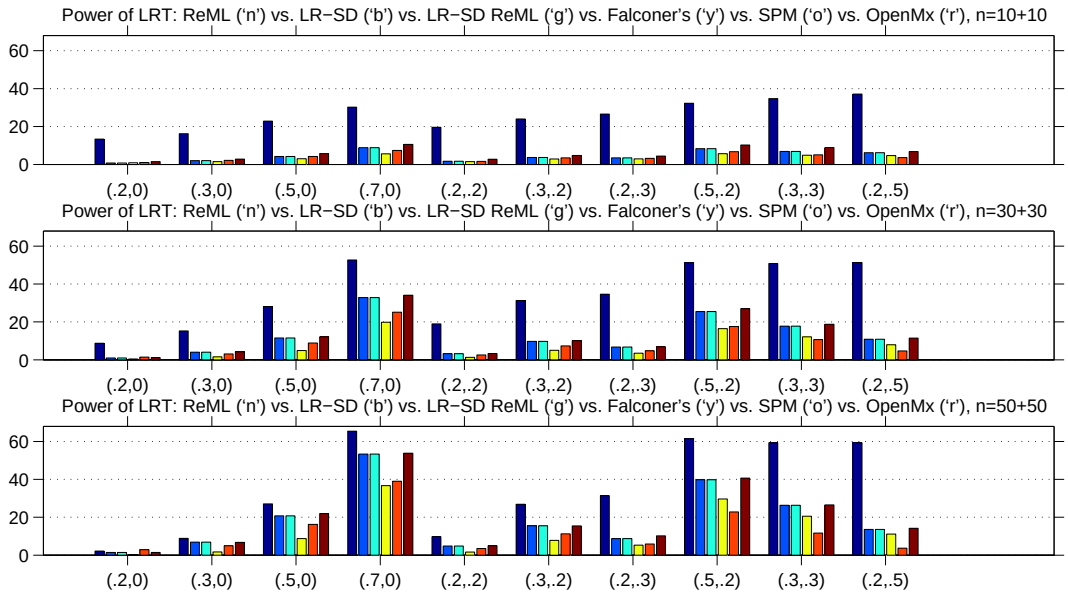


Figure 3.5: The comparison of statistical power of LRT in percent at $\alpha = 0.05$ with false null hypothesis ($H_0 : h^2 = 0$): Frequentist ReML ('n': navy), LR-SD ('b': blue), LR-SD ReML ('g': green), Falconer's method ('y': yellow), Bayesian ReML used in SPM ('o': orange), and SEM used in OpenMx ('r': red). Comma ordered pairs on x-axis correspond to the rounded parameter values of (A, C) with $A > 0$; see Table 3.1 and Figure 3.3 for exact parameter settings used.

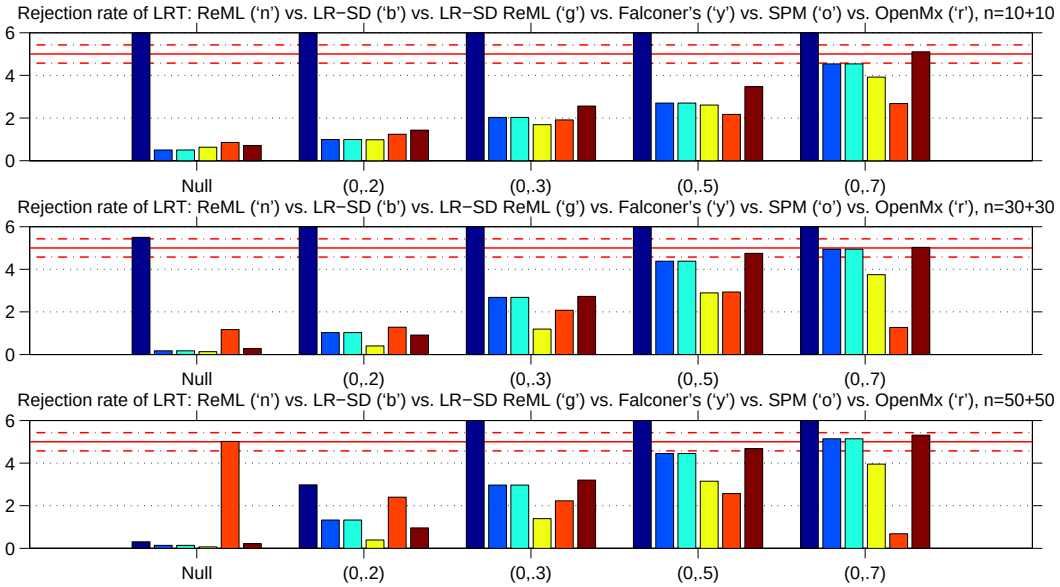


Figure 3.6: The comparison of the estimated FPR (false rejection rate) of LRT in percent at level $\alpha = 0.05$ with true null hypothesis ($\mathrm{H}_0 : h^2 = 0$): Frequentist ReML ('n': navy), LR-SD ('b': blue), LR-SD ReML ('g': green), Falconer's method ('y': yellow), Bayesian ReML used in SPM ('o': orange), and SEM used in OpenMx ('r': red). Comma ordered pairs on x-axis correspond to the rounded parameter values of (A, C) with $A = 0$; see Table 3.1 and Figure 3.3 for exact parameter settings used. The two red dash-dotted lines show the lower and upper bounds of the 95% binomial proportion confidence interval. The FPR should be 0.05, but its estimates can vary within the 95% binomial proportion confidence interval $[0.0457, 0.0543]$ for $n_{\mathrm{Rlz}} = 10000$ simulations.

dash-dotted lines) indicates valid tests are used.

Running Time The running time of entirely $n_{\mathrm{Rlz}} = 10000$ simulations for these 6 variance components estimation methods is shown in Figure 3.7. The computational performance comparing all these methods reveals Falconer's method and the linear regression methods with SD's—LR-SD and LR-SD ReML always outperform other iterative methods including Frequentist ReML, Bayesian ReML in SPM and SEM in OpenMx. For each simulation setting, the overall computation time of all simulations for those non-iterative methods is far smaller than the other iterative methods. Specifically, LR-SD is roughly 2.5 times faster than LR-SD ReML and around 300 times faster than SEM in OpenMx on average, which indicates that an evident decrease of the computation time has been achieved. However, the compu-

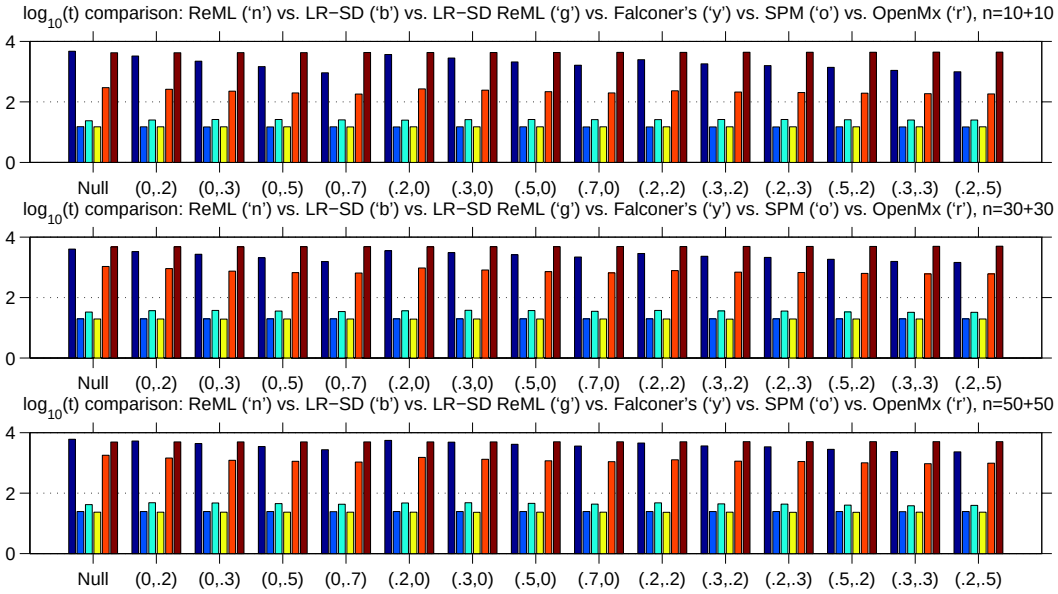


Figure 3.7: The total running time comparison for $n_{\text{Rlz}} = 10000$ simulations after base-10 log transformation ($\log_{10}(t)$): Frequentist ReML ('n': navy), LR-SD ('b': blue), LR-SD ReML ('g': green), Falconer's method ('y': yellow), Bayesian ReML used in SPM ('o': orange), and SEM used in OpenMx ('r': red). Comma ordered pairs on x-axis correspond to the rounded parameter values of (A, C); see Table 3.1 and Figure 3.3 for exact parameter settings used.

tation time of Frequentist ReML is prohibitive with the running time of around 4 times longer than that of Bayesian ReML in SPM averagely.

3.5.2 ROC-based Power Evaluation

To compare the voxel- and cluster-wise heritability inference approaches described in Section 2.5.2, receiver operating characteristic (ROC) curves are applied, with the use of our newly proposed univariate LR-SD method, to examine the statistical sensitivity of these two approaches.

Simulation Setting

With simulations, we now intend to evaluate the statistical power (sensitivity) of the voxel- and cluster-wise inference approaches using our LR-SD method for univariate heritability analysis. The sample sizes considered are exactly the same as those mentioned above in 1D simulations to be $n = 20, 60, 100$, where only twins are

involved in the analysis [5] and the number of MZ and DZ twins is equal. The signal data is generated with 6 parameter settings shown in Table 3.2 consisting of different extents of heritability and environmental sharing. For the null heritability settings, varying degrees of shared environmental effect are investigated including $C = 0, \frac{1}{3}, \frac{2}{3}$. While for the non-null heritability settings, there are 3 levels of heritability settings considered of $h^2 = \frac{1}{3}, \frac{1}{2}, \frac{2}{3}$. Since our main focus is on the heritability, we only set one setting with non-zero C effect. For this non-null setting, the unique environmental effect E is set the same as another non-null setting so as to examine the influence of the C effect by comparing these two settings with identical E effect. Without loss of generality, the parameters A, C and E are also scaled so that unit phenotypic variance $\sigma^2 = 1$ is used.

Table 3.2: 6 parameter settings of heritability h^2 and parameters $(A, C, E)^{\mathrm{T}}$.

	h^2	A	C	E
No Heritability	0	0	0	1
($h^2 = 0$)	0	0	1/3	2/3
	0	0	2/3	1/3
Positive Heritability	1/3	1/3	0	2/3
($h^2 > 0$)	1/2	1/2	1/6	1/3
	2/3	2/3	0	1/3

Without loss of generality, the simulated images are 2D with the image size of 128×128 pixels. The process of ground noise image generation starts by simulating multiple 2D i.i.d. Gaussian random images; the heritability signal images are created later by substituting the signal-specific part within the generated ground noise images with the simulated signal data. A range of spatial Gaussian smoothing kernels with full width at half maximum (FWHM) of {0, 1.5, 3, 6} pixels are applied for both noise and signal images separately to take into account of the similarity of the neighboring voxels. We consider two tested signal shapes of the focal and distributed that are shown in Figure 3.8 with matched number of involving signal pixels. Totally $n_{\mathrm{Rlz}} = 1000$ images were generated for each simulation setting.

[5] During 1D simulation, we found adding singletons can improve neither estimation accuracy nor statistical power. However, we still suggest including singletons in the real data analysis since a better estimate of the phenotypic variance can be obtained with more data taken into account.

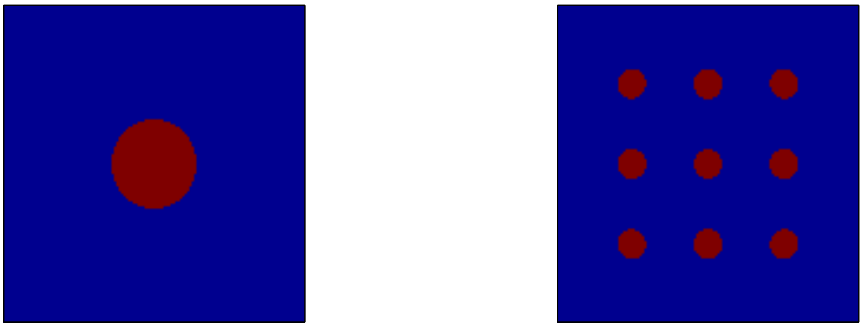


Figure 3.8: Illustration of the 2D simulated signal shapes. Focal signal (left) with 1 large circle in the middle; distributed signal (right) with 9 identical small circles.

Power Analysis

The ROC curves plot y axis as the true positive rate (TPR) against x axis as FPR with varying threshold levels. In practice, the statistical measures of sensitivity (or TPR) and specificity (or $1 - \text{FPR}$) can not be both large, so the ROC curves reflect the trade-off between them. An ideal inference method will give 100% sensitivity (i.e., TPR $= 1$) and 100% specificity (i.e., FPR $= 0$), which yield a point $(0, 1)$ in the ROC curve, and TPR will stay at 1 for all other FPR values (Smith and Nichols, 2009). As each 2D image contains multiple pixels and the standard ROC method was designed for a single test, the alternative free-response ROC approach (Chakraborty and Winter, 1990), which uses the fraction of images with any false positives as x axis and the proportion of true positives detected as y axis, is employed to control the family-wise error rate in multiple comparisons problem (Smith and Nichols, 2009).

A commonly-used summary measure of the ROC curves, called the area under the curve (AUC), is applied to summarize the ROC curves with larger value indicating higher statistical power (e.g., AUC $= 1$ for the perfect methods). Since we are more concerned about FPR values between 0 and 0.05 when the significance level of a test is set $\alpha = 0.05$, the normalized AUC, expressed as $20 \times$ AUC for FPR $= 0 : 0.05$, is calculated for both voxel- and cluster-wise inference approaches for statistical power comparison.

The following steps clarify this ROC-based approach for power evaluation of voxel- and cluster-wise inference approaches for each simulation setting.

(i) Generate $n_{\text{Rlz}} = 1000$ i.i.d. 2D Gaussian random noise images and the corresponding $n_{\text{Rlz}} = 1000$ heritability signal images.

(ii) For each image, estimate heritability pixel-by-pixel and create the test statistic image.

(iii) **Voxel-wise Inference** Apply a large number of pre-defined thresholds, obtain the suprathreshold pixels, and then calculate family-wise FPR and TPR for each of these threshold levels.

Cluster-wise Inference Threshold the LRT statistic images with an arbitrary pre-determined cluster-forming threshold (e.g., 0.05) and form suprathreshold clusters for all the statistic images. For each observed suprathreshold cluster, consider a full range of possible cluster sizes as the thresholds, and then compute family-wise FPR and TPR for each threshold level (i.e., cluster size).

- **Family-wise FPR Computation** The ground noise images are used for FPR calculation for each threshold. For each noise image, count the number of images with one or more pixels detected, which is divided by the number of realizations (i.e., $n_{\text{Rlz}} = 1000$) to obtain an estimate of the family-wise FPR.
- **TPR Computation** The heritability signal images are used for TPR calculation for each threshold. For each signal image, count the number of true heritability pixels detected ($\#\{\text{TPs}\}$), which is divided by the total number of true heritability pixels ($\#\{\text{TPs} + \text{FNs}\}$) to get a proportion of true positive detection (r_{TPs}), i.e.,

$$r_{\text{TPs}} = \frac{\#\{\text{TPs}\}}{\#\{\text{TPs} + \text{FNs}\}},$$

where "TPs" and "FNs" denote true positive and false negative decisions respectively [6]. Average this proportion over all realizations to derive an estimate of TPR.

(iv) Plot the ROC curves and calculate the corresponding normalized AUC values.

ROC-based Simulation Results

As described above, a range of simulation settings are investigated for both voxel- and cluster-wise inference approaches using LR-SD. For different extents of smooth-

[6] Apart from this true positive proportion r_{TPs}, a so-called "rand index" measure or its adjusted form can also be utilized to compare the true data clustering and our detected clustering results, and to evaluate the similarity between them, i.e., the accuracy of the search results.

ness, the returned ROC curves have fairly similar shape, so we will only illustrate the ROC curves created by medium degree of smoothing with FWHM of 3 pixels, which are shown in Figures 3.9 and 3.10 for the simulated focal and distributed signals respectively, and leave out those by other smoothness extents. The corresponding normalized AUC comparison is then shown in Figure 3.11.

For the focal signal, the results are shown in Figure 3.9. The ROC curves of voxel-wise method are always below those of cluster-wise method for different sample sizes and all parameter settings, which reveals a higher statistical power obtained for cluster- than voxel-wise inference approaches. For a particular family-wise FPR level, the TPR value of both inference methods becomes larger when the sample size is enlarged or the heritability extent is increased. Compared with the tightly focal signal, the spatially distributed signal generates ROC curves, shown in Figure 3.10, behaving analogically in nearly all respects to the focal signal.

In Figure 3.11, we summarize the above ROC curves and follow up with a bar plot of the normalized AUC, which is a meaningful ROC-related summary measure. As stated above, we only care about FPR between 0 and 0.05 (i.e., FPR less than the level $\alpha = 0.05$), and the normalized AUC for FPR $= 0 : 0.05$ is employed and calculated rather than the standard AUC. The voxel-wise method has poor performance overall for all simulation settings with negligible AUC values, while the cluster-wise inference approach has much larger AUC values, suggesting that the cluster-wise method is more statistically powerful and motivating the use of the cluster-wise approach.

3.6 Real Data Applications

In this section, we apply the voxel- and cluster-wise heritability inference approaches using univariate LR-SD method to two fMRI datasets, with one investigating the working memory brain activation and another associated with the amygdala-related emotional stimuli.

3.6.1 Heritability of Working Memory Brain Activation

This section contributes to the investigation of the heritability of working memory task-related brain activation. We illustrate the above-mentioned heritability inference approaches including univariate LR-SD and permutations.

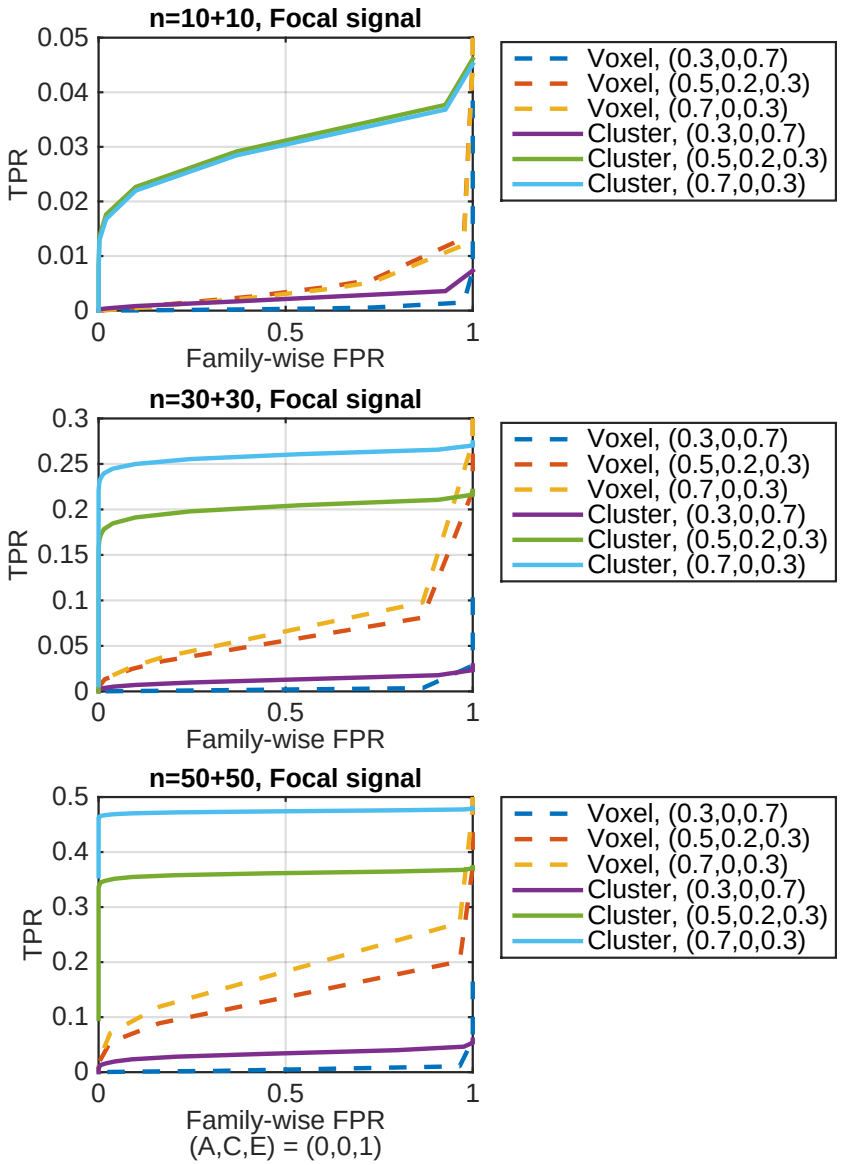


Figure 3.9: The ROC curve comparison of voxel- (dashed lines) and cluster-wise (solid lines) inference approaches for different parameter settings of $(A, C, E)^{\mathrm{T}}$ for the focal signal with 3 sample sizes of 10+10 (upper), 30+30 (middle) and 50+50 (lower).

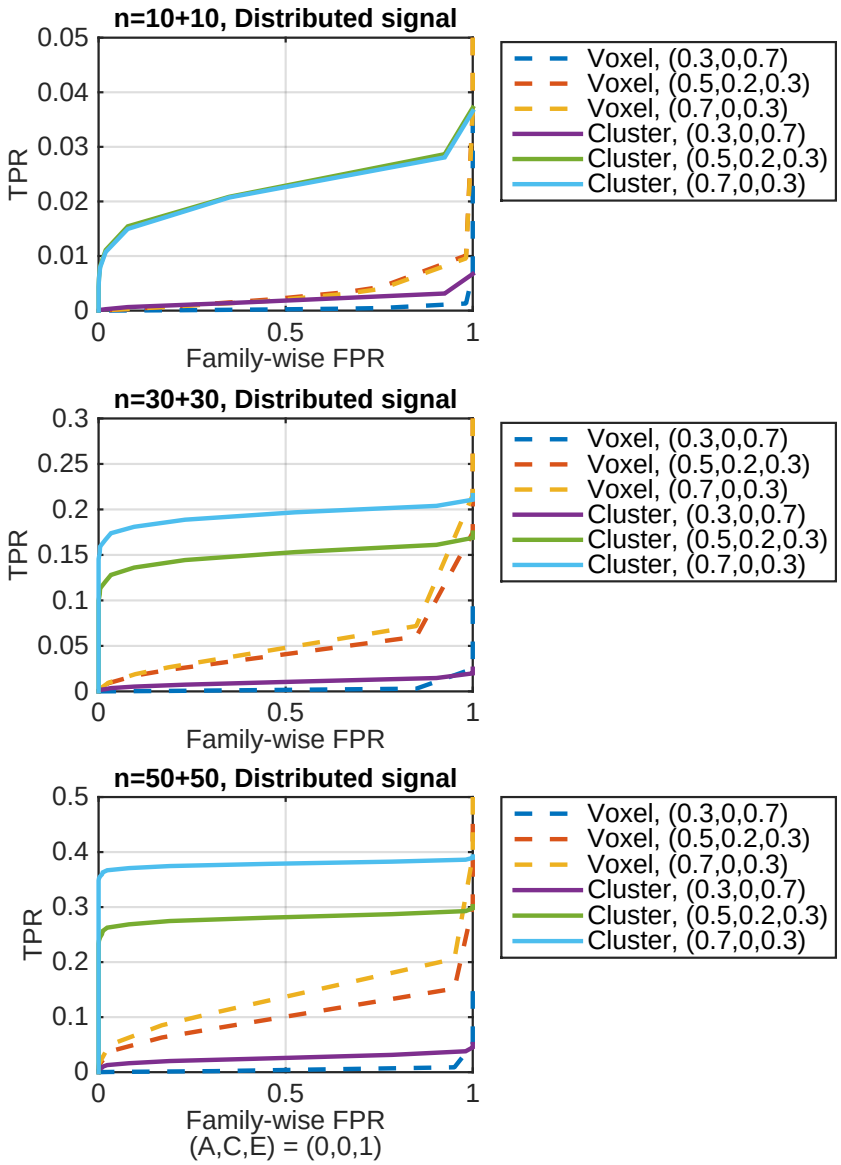


Figure 3.10: The ROC curve comparison of voxel- (dashed lines) and cluster-wise (solid lines) inference approaches for different parameter settings of $(A, C, E)^{\mathrm{T}}$ for the distributed signal with 3 sample sizes of 10+10 (upper), 30+30 (middle) and 50+50 (lower).

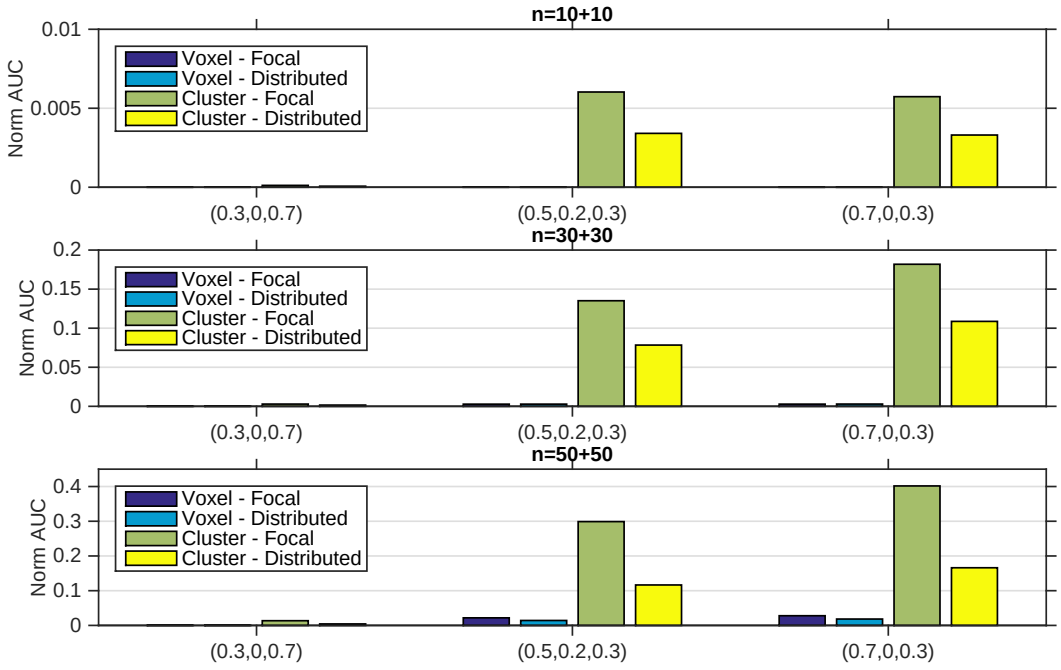


Figure 3.11: The normalized AUC ($20 \times$ AUC for FPR $= 0 : 0.05$) comparison of voxel- and cluster-wise inference approaches for different parameter settings of $(A, C, E)^{\mathrm{T}}$, 3 samples of size $n = 10 + 10, 30 + 30, 50 + 50$, and two tested signals (focal and distributed) with positive heritability $h^2 > 0$.

Real Data Acquisition

The experimental sample comprises entirely $n = 319$ young and healthy participants (199 females and 120 males), consisting of $n_{\mathrm{MZ}} = 150$ MZ twins (75 pairs with 46 female and 29 male), $n_{\mathrm{DZ}} = 132$ DZ twins (66 pairs with 30 female, 11 male and 25 opposite sex) and $n_{\mathrm{S}} = 37$ unpaired twins (22 female and 15 male). The age range of all these subjects is 20–28 years (mean $\pm$ SD : 23.6 ± 1.8). They were invited to join the study in Brisbane, Australia, and performed the 0-back and 2-back working memory tasks during the experiment. The 4T Bruker Medspec full-body scanner was utilized and task-related fMRI BOLD signals were acquired to create the brain images. Imaging pre-processing was implemented using the SPM5 software in Matlab, including image realignment with a mean image generated, spatial normalization to the standard T1 template in MNI atlas space, spatial smoothing with isotropic Gaussian kernel, removal of global signal effects, and the use of high-pass and low-pass filtering to discard uninterested signals. For each subject, the brain activation, measured as the 2-back $>$ 0-back t-contrast images using one-sample t test, was extracted. Only areas of expected activation are included in the mask of $K = 14627$ in-mask voxels in total. Age, gender and 2-back performance accuracy (the percentage of correct responses) are considered and included as the covariates in the statistical analysis (Blokland et al., 2011).

Results

As mentioned previously, normally there are quite massive possible permutations in total, which requires large amounts of computation time. Since an effective and efficient approximation to the permutation distribution can be pursued with a large number of permutations, the characteristics of the permutation distribution can be acquired with fewer permutations (Nichols and Holmes, 2001). Here we perform N = 1000 permutations as suggested. One run of Mx took around two days on this fMRI dataset while LR-SD only took about 6 minutes on a MacBook Pro with dual quad-core CPUs (3.0 GHz). Applying our LR-SD method voxel-by-voxel and making permutation inference on the fMRI twin data, all these N = 1000 permutations took around 15.5 hours by separating them into 10 parallelized jobs, each with 100 permutations.

The permutation-based empirical distributions of maximum test statistics are shown in Figure 3.12 for maximum LRT statistic $\mathrm{T}_{\mathrm{p}}^{\max}$, in Figure 3.13 for maximum suprathreshold cluster size $\mathrm{K}_{\mathrm{p}}^{\max}$, and in Figure 3.14 for maximum suprathreshold cluster

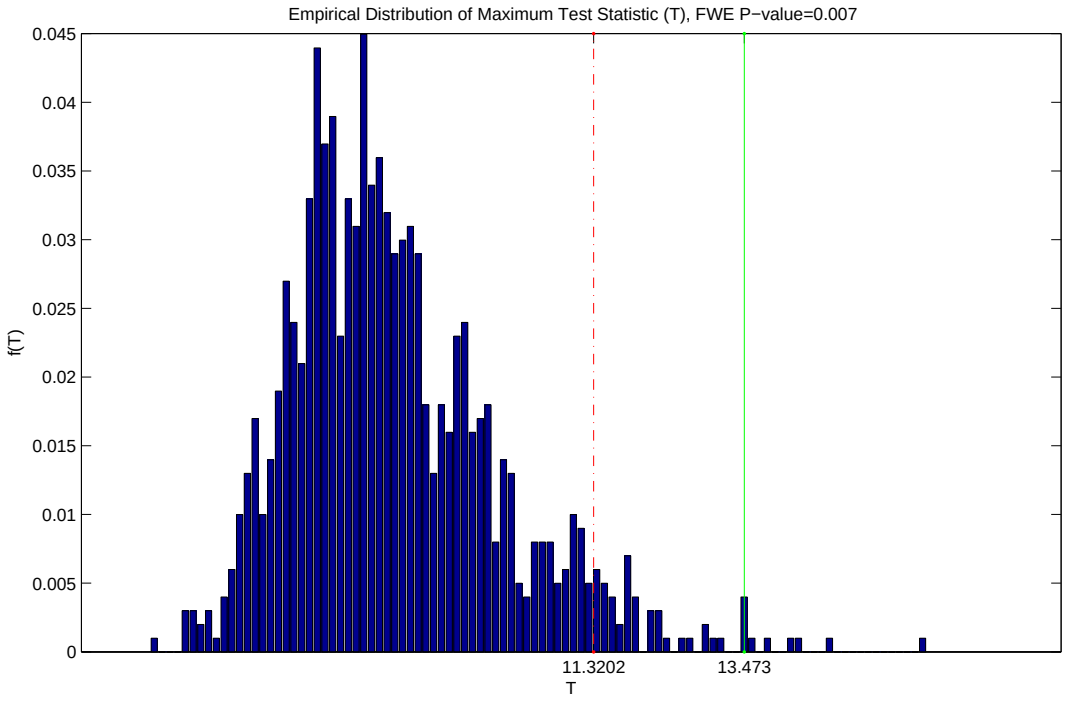


Figure 3.12: Empirical permutation distribution of maximum LRT statistic T_p^{max} with red dash-dot line for the critical threshold at level $\alpha = 0.05$ and green solid line for the observed maximum LRT statistic value.

mass M_p^{max}. The appointed significance level is chosen to be $\alpha = 0.05$, leading to the critical threshold to be the 51^{th} largest member of each permutation distribution, with $T_\alpha = 11.3202$, $K_\alpha = 62$ and $M_\alpha = 271.7397$. The most significant FWE-corrected p-values are 0.007, 0.001 and 0.001 for voxel in Figure 3.12, for cluster size in Figure 3.13, and for cluster mass in Figure 3.14 respectively, which implies that the omnibus hypothesis of no heritability everywhere in the ROI's is rejected in the single threshold test (see Figure 3.12) and suprathreshold cluster tests (see Figures 3.13 and 3.14) at level $\alpha = 0.05$.

The FWE-corrected p-value images after log transformation (i.e., $-\log_{10}(p_{\text{FWE}})$) for significant voxels and significant suprathreshold clusters with respect to size and mass statistics are shown in Figure 3.15, and Figures 3.16 and 3.17, respectively. Suprathreshold cluster tests (see Figures 3.16 and 3.17) found much larger significant brain areas than single threshold test (see Figure 3.15) by comparing the FWE-corrected p-value images. Only 2 significant voxels are found for voxel-wise single

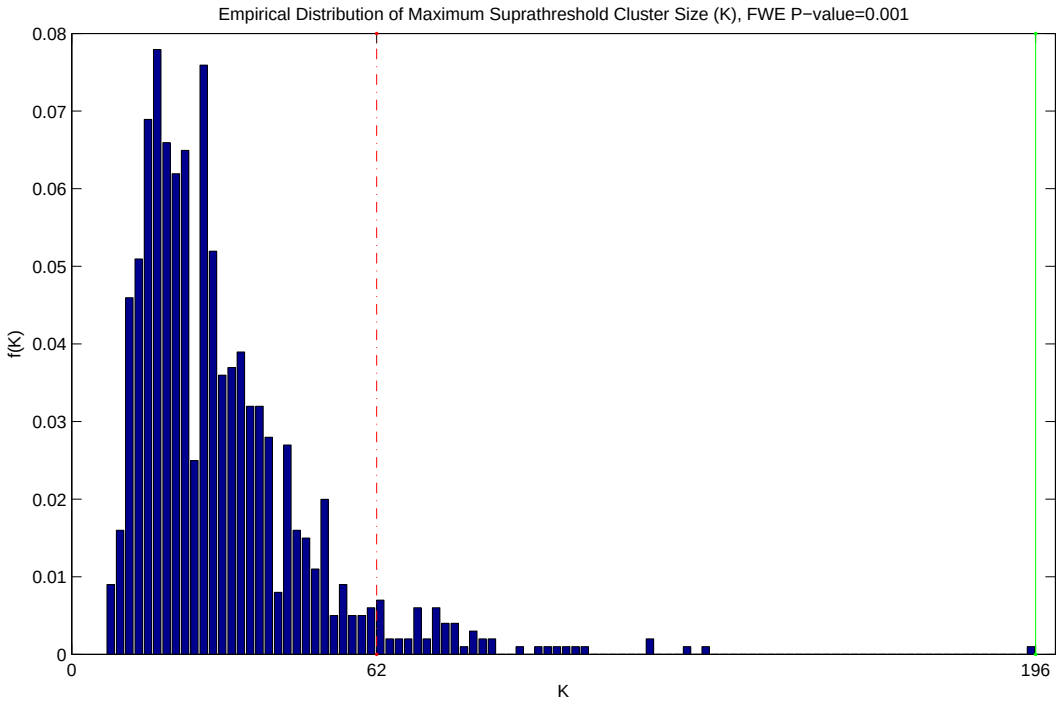


Figure 3.13: Empirical permutation distribution of maximum suprathreshold cluster size K_p^{max} with red dash-dot line for the critical threshold at level $\alpha = 0.05$ and green solid line for the observed maximum suprathreshold cluster size.

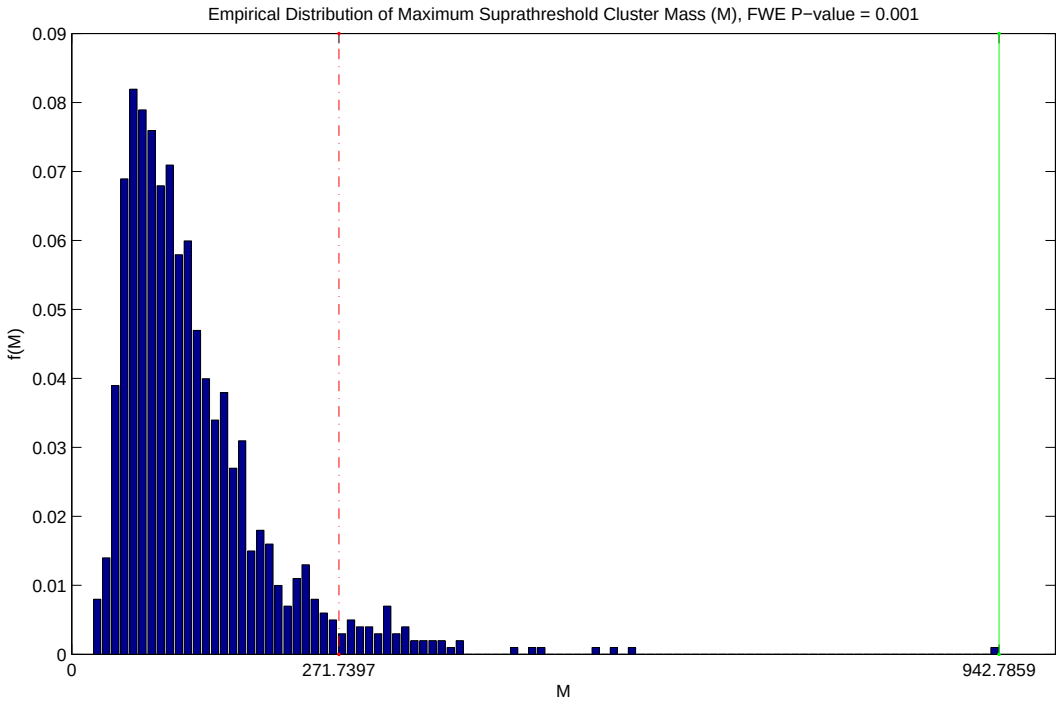


Figure 3.14: Empirical permutation distribution of maximum suprathreshold cluster mass M_p^{max} with red dash-dot line for the critical threshold at level $\alpha = 0.05$ and green solid line for the observed maximum suprathreshold cluster mass.

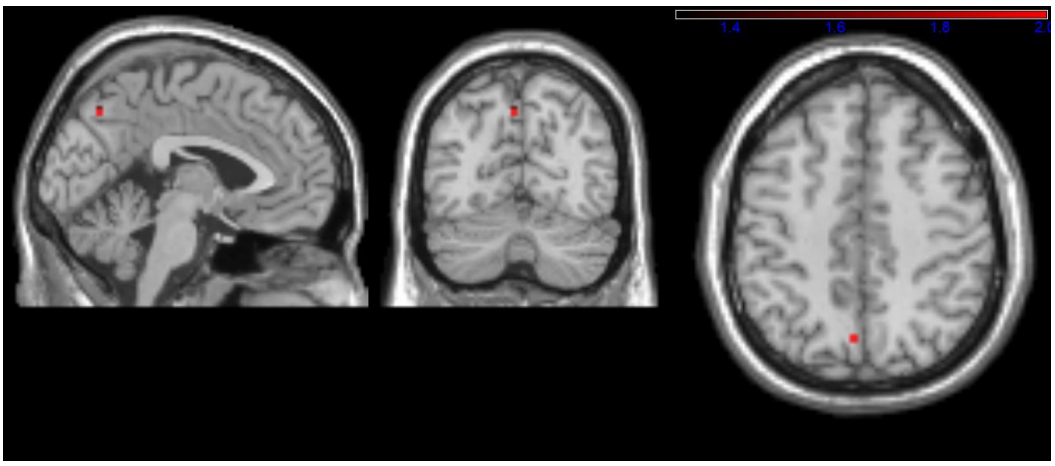


Figure 3.15: The log transformed p-value image (i.e., $-\log_{10}(p_{\text{FWE}})$) for voxels with significant LRT statistic.

test, while 4 significant clusters with a total of 634 voxels are found for cluster-wise tests. All these heritability-significant regions found in both single threshold test and suprathreshold cluster tests are the memory-related brain regions, which is consistent with the previous findings (Blokland et al., 2011).

3.6.2 Heritability of Amygdala Response to Emotional Stimuli

In this section, the heritability in response to the emotional stimuli is analyzed using summary statistics of the ACE model over the brain region of amygdala, which is a brain region typically associated with emotional processing. We provide inference on the unweighted mean summaries for illustration, which is compared with the results obtained from voxel- and cluster-wise inferences using the maximum statistics.

Real Data Acquisition

A sample of $n = 111$ subjects in total from a pre-processed fMRI dataset [7] includes $n_{\text{MZ}} = 32$ MZ twins (16 pairs), $n_{\text{DZ}} = 50$ DZ twins (25 pairs) and $n_{\text{S}} = 29$ singletons [8]. These 111 subjects are all males, aged 10–12, from the Twins Early Development

[7] The access information of this fMRI dataset was provided in an email from Prof. T.E. Nichols (t.e.nichols@warwick.ac.uk) in July 2011.

[8] Although we found the statistical power of the effect size $h^2 = 0.5$ for this small sample is low (less than 20%) using simulations, and nothing significant was found using the voxel-wise inference for this dataset, the use of spatial statistics derived from cluster and summary inferences increases the power of the test at the same significance level.

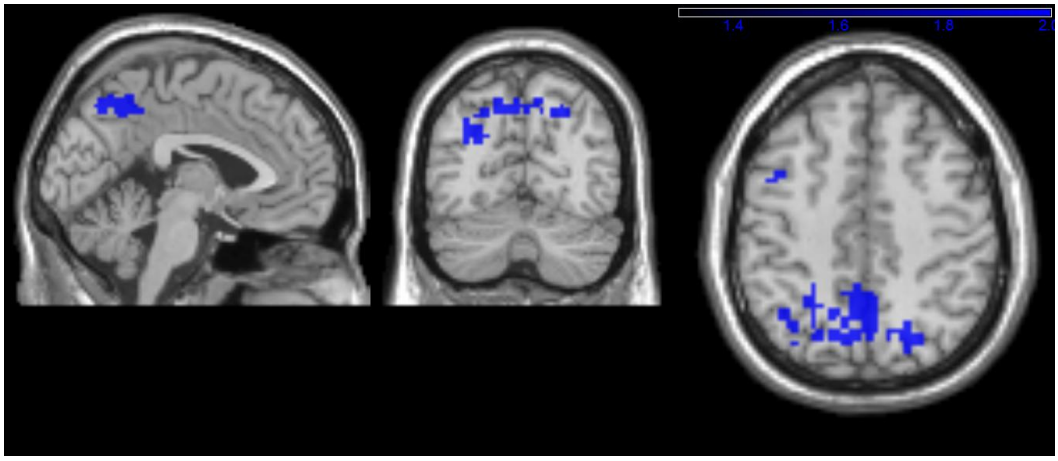


Figure 3.16: The log transformed p-value image (i.e., $-\log_{10}(p_{\text{FWE}})$) for suprathreshold clusters on the observed image with significant suprathreshold cluster size.

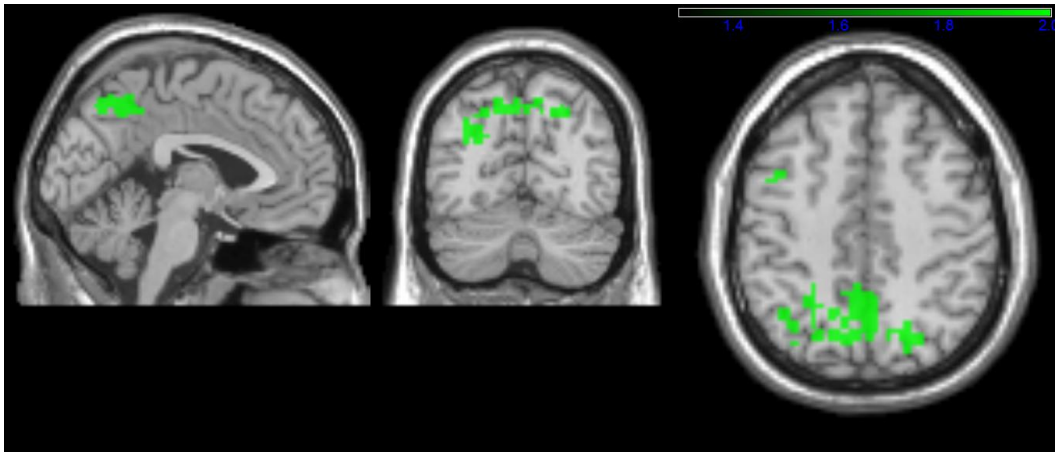


Figure 3.17: The log transformed p-value image (i.e., $-\log_{10}(p_{\text{FWE}})$) for suprathreshold clusters on the observed image with significant suprathreshold cluster mass.

Study (TEDS) [9]. There are 50 participants out of 111 with the behavioural problems by SDQ assessment. Hence these 111 participants can be separated into two groups: subjects with behavioural problems (CPCU+), and healthy subjects (Negative). During the experiment, all subjects equally performed an emotional pictures matching task with the pictures from International Affective Picture System (IAPS) (Coan and Allen, 2007). The brain ROI's considered are left amygdala, right amygdala and both amygdalas of overall 286, 269 and 555 in-mask voxels separatively. There are 3 samples including all subjects, CPCU+ ($n_+ = 50$ with 14 MZ twins, 30 DZ twins and 6 singletons) and Negative ($n_- = 61$ with 18 MZ twins, 20 DZ twins and 23 singletons) considered for heritability analysis. Both permutation and bootstrapping inferences are utilized to compute the p-values and confidence intervals. Before making the analysis, the inverse Gaussian transformation was implemented with the use of probability integral transform to assure the normality of the response data and remove the possible outliers, which is routinely designed to make the response data variable with a skewed or non-normal distribution follow a normally distributed pattern.

Results

In Tables 3.3 and 3.4, the results for the three samples of all subjects, the CPCU+ group and the Negative group are shown. For all subjects, the estimates, 95% bootstrapping confidence intervals from 1000 bootstrap replicates, and permutation-based p-values derived using 1000 permutations for different spatially informed test statistics are explicitly listed for the considered brain ROI's of both amygdalas, left amygdala and right amygdala in Table 3.3. The spatial statistics employed include the unweighted mean summaries of heritability ($\overline{h^2}$) and common environmental factor ($\overline{c^2} = \frac{1}{K}\sum_{r=1}^{K} c_r^2$ for all K in-mask voxels, where c_r^2 denotes the common environmental factor for voxel r ($r = 1, \ldots, K$)), and the maximum statistics of maximum LRT statistic ($\mathrm{T_p^{max}}$) from voxel-wise inference, and maximum suprathreshold cluster size ($\mathrm{K_p^{max}}$) and cluster mass ($\mathrm{M_p^{max}}$) from cluster-wise inference. For CPCU+ and Negative, the estimates and permutation-based p-values derived using 1000 permutation for the above-mentioned spatial statistics are displayed in Table 3.4 for those 3 amygdala ROI's. In Table 3.4, the difference between these two groups is also shown in terms of the estimate and its 95% bootstrapping confidence interval from 1000 bootstrap replicates.

For all subjects, the cluster-wise statistics of maximum suprathreshold cluster size

[9] `http://www.teds.ac.uk/`

and cluster mass and the unweighted mean statistic of heritability are found to be significant at level $\alpha = 0.05$ for the 3 considered brain ROI's, but the voxel-wise statistic (maximum LRT statistic) is only found to be significant for left amygdala. Compared with right amygdala, left amygdala has a lower heritability summary measure ($\overline{h^2} = 0.4108$ for left amygdala vs. $\overline{h^2} = 0.4564$ for right amygdala), but its maximum LRT statistic value is higher than that of right amygdala ($T_p^{max} = 6.3357$ for left amygdala vs. $T_p^{max} = 5.7304$ for right amygdala) although its cluster size statistic is smaller ($K_p^{max} = 70$ for left amygdala vs. $K_p^{max} = 97$ for right amygdala), which reveals that left amygdala has a higher peak voxel while right amygdala has a larger significant cluster.

In Table 3.4, we only find significance using mean heritability summary statistics for those included brain regions for the CPCU+ group, while the evidence for heritability is found with cluster statistics for the Negative group for right amygdala and both amygdalas, but not for left amygdala. There is no significance found using voxel-wise statistic for either CPCU+ or Negative, which should be accounted for by the reduced statistical power for small samples. Comparing the CPCU+ and Negative groups using their difference, we find that the mean heritability measure of CPCU+ is always smaller than that of Negative, and the values of spatial statistics including the voxel-wise peak height and cluster size and mass for CPCU+ are much smaller than those for Negative, indicating that these two groups are distinct regarding heritability on amygdala.

Even for the larger sample of all subjects in Table 3.3, the 95% bootstrapping confidence intervals for $\overline{c^2}$ still contain the tested value of 0, implying that the omnibus hypothesis test of zero common environmental factor cannot be rejected. In Table 3.4, even if the estimates of heritability mean summary $\overline{h^2}$ for CPCU+ are less than those for Negative for all 3 samples, $\overline{h^2}$ is always found significant for CPCU+ while insignificant for Negative, which may be due to the smaller mean measure of common environmental factor $\overline{c^2}$ for CPCU+.

3.7 Summary of the Chapter

In this chapter, we have presented 3 novel voxel-wise heritability estimation methods trying to improve the estimation accuracy and statistical sensitivity with a controlled false positive rate (specificity), and reduce the computational complexity. The simple LR-SD method based on linear regression modeling with squared differ-

ences of paired observations has been introduced and established. This method is found to have comparable or even better estimation accuracy and statistical power relative to the existing methods and the other two newly proposed methods. Simulation studies also show that LR-SD is the most computationally efficient approach overall and will never encounter any convergence problems compared with the other iterative methods. These time-efficient, accurate and non-iterative properties of LR-SD make it more flexible and feasible to be applied for permutation inference to correct for the family-wise error rate and bootstrapping inference to construct the confidence intervals.

Table 3.3: For all $n = 111$ subjects, the estimates, 95% bootstrapping confidence intervals from 1000 bootstrap replicates ("95% CI") and permutation-based p-values derived using 1000 permutations for the unweighted mean summaries of $\overline{h^2}$ for heritability and $\overline{c^2}$ for shared environmental factor, and the maximum statistics of T_p^{max}, K_p^{max} and M_p^{max} from the voxel- and cluster-wise inferences are obtained for 3 brain ROI's including both amygdalas, left amygdala and right amygdala.

	Both Amygdalas		
	Estimate	95% CI	P-value
$\overline{h^2}$	0.4329	(0.2145, 0.6012)	0.003
$\overline{c^2}$	0.0037	(0.0000, 0.1682)	/
T_p^{max}	6.3357	/	0.125
K_p^{max}	97	/	0.017
M_p^{max}	360.264	/	0.026
	Left Amygdala		
	Estimate	95% CI	P-value
$\overline{h^2}$	0.4108	(0.1993, 0.6093)	0.005
$\overline{c^2}$	0.0019	(0.0000, 0.1421)	/
T_p^{max}	6.3357	/	0.037
K_p^{max}	70	/	0.006
M_p^{max}	261.017	/	0.008
	Right Amygdala		
	Estimate	95% CI	P-value
$\overline{h^2}$	0.4564	(0.2218, 0.6367)	0.004
$\overline{c^2}$	0.0056	(0.0000, 0.2095)	/
T_p^{max}	5.7304	/	0.151
K_p^{max}	97	/	0.023
M_p^{max}	360.264	/	0.032

Table 3.4: For both groups of CPCU+ (denoted as "+" and with $n_+ = 50$ subjects) and Negative (denoted as "–" and with $n_- = 61$ subjects), the estimates ("Est") and permutation-based p-values derived using 1000 permutations ("P") for the unweighted mean summaries of $\overline{h^2}$ for heritability and $\overline{c^2}$ for shared environmental factor, and the maximum statistics of T_p^{max}, K_p^{max} and M_p^{max} from the voxel- and cluster-wise inferences are obtained for 3 brain ROI's including both amygdalas, left amygdala and right amygdala. The estimates and 95% bootstrapping confidence intervals from 1000 bootstrap replicates ("95% CI") for the difference between these two groups are also shown in the table below.

	Both Amygdalas					
	Est ("+")	P ("+")	Est ("–")	P ("–")	Est (Diff)	95% CI (Diff)
$\overline{h^2}$	0.3100	0.024	0.5029	0.068	0.1929	(0.0000, 0.6518)
$\overline{c^2}$	0.0078	/	0.0810	/	0.0732	(0.0000, 0.5197)
T_p^{max}	3.5229	0.329	6.6650	0.134	/	/
K_p^{max}	3	0.333	91	0.025	/	/
M_p^{max}	9.543	0.311	343.952	0.024	/	/
	Left Amygdala					
	Est ("+")	P ("+")	Est ("–")	P ("–")	Est (Diff)	95% CI (Diff)
$\overline{h^2}$	0.2883	0.022	0.4993	0.120	0.2110	(0.0000, 0.6621)
$\overline{c^2}$	0.0017	/	0.0553	/	0.0536	(0.0000, 0.5394)
T_p^{max}	2.8162	0.217	6.3375	0.150	/	/
K_p^{max}	2	0.174	61	0.065	/	/
M_p^{max}	5.618	0.174	206.803	0.075	/	/
	Right Amygdala					
	Est ("+")	P ("+")	Est ("–")	P ("–")	Est (Diff)	95% CI (Diff)
$\overline{h^2}$	0.3315	0.043	0.5067	0.059	0.1752	(0.0000, 0.6651)
$\overline{c^2}$	0.0139	/	0.1087	/	0.0948	(0.0000, 0.4813)
T_p^{max}	3.5229	0.274	6.6650	0.063	/	/
K_p^{max}	3	0.272	91	0.020	/	/
M_p^{max}	9.543	0.247	343.952	0.020	/	/

Chapter 4

Multivariate Genetic Inference

In this chapter, we extend the univariate heritability inference approach discussed in Section 3.4 to the multivariate case. The multivariate general linear model accounting for multiple phenotypes or a high-dimensional phenotype is established, the unknown phenotypic and genetic correlations underlying the phenotypic traits are estimated, and then the hypothesis of zero genetic correlation is tested. We investigate two multivariate inference approaches including bivariate LR-SD, the generalization of univariate LR-SD described in Section 3.4, and correlation mean difference to analyze the genetic effects. During the analysis, we only take twins and singletons into consideration.

4.1 The General Linear Model

Assume that there are J phenotypes (or J phenotypic elements for a high-dimensional phenotype) considered for the multivariate modeling. Suppose the sample contains totally n subjects, as stated in Section 3.1, including n_{MZ} MZ twins ($\frac{1}{2}n_{\mathrm{MZ}}$ pairs), n_{DZ} DZ twins ($\frac{1}{2}n_{\mathrm{DZ}}$ pairs) and n_{S} singletons. In this chapter, we require an additional index k to account for the different phenotypes jointly modeled. For a particular voxel, the voxel-wise data vector for phenotype k is denoted as $\mathbf{Y}^{(k)}$ $(k = 1, \ldots, J)$ by eliminating the voxel index r. We suppose the number of covariates included for the phenotype k is $p_k - 1$. Generalizing the univariate GLM model (3.1) for a single phenotype with the data vector $\mathbf{Y}^{(k)}$ to the multivariate case yields

the multivariate GLM with respect to all these J phenotypes:

$$\begin{pmatrix} \mathbf{Y}^{(1)} \\ \vdots \\ \mathbf{Y}^{(J)} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^{(1)}\boldsymbol{\beta}^{(1)} + \boldsymbol{\epsilon}^{(1)} \\ \vdots \\ \mathbf{X}^{(J)}\boldsymbol{\beta}^{(J)} + \boldsymbol{\epsilon}^{(J)} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{X}^{(1)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{X}^{(J)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^{(1)} \\ \vdots \\ \boldsymbol{\beta}^{(J)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}^{(1)} \\ \vdots \\ \boldsymbol{\epsilon}^{(J)} \end{pmatrix},$$

which can be simply written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{4.1}$$

where $\mathbf{Y} = (\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(J)})^{\mathrm{T}}$ is an nJ-vector of nJ observations, the design matrix $\mathbf{X}$ is a block-diagonal design matrix with blocks specified as $\mathbf{X}^{(k)}$ $(k = 1, \ldots, J)$, which is corresponding to phenotype k, $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(J)})^{\mathrm{T}}$ is a column vector of $\sum_{k=1}^{J} p_k$ regression coefficients, and the error vector $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(J)})^{\mathrm{T}}$ is assumed to be normally distributed with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{V}$. The covariance matrix $\mathbf{V}$ has the structure of

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}^{(1)} & \cdots & \mathbf{C}^{(1J)} \\ \vdots & \ddots & \vdots \\ \mathbf{C}^{(J1)} & \cdots & \mathbf{V}^{(J)} \end{pmatrix},$$

where each diagonal block $\mathbf{V}^{(k)}$ is the phenotypic variance matrix associated with phenotype k, and the off-diagonal covariance matrix $\mathbf{C}^{(kk')}$ for a pair of phenotypes k and k' $(k \neq k'$ and $k, k' = 1, \ldots, J)$ can be analogously constructed with a similar structure as $\mathbf{V}^{(k)}$ (see details below). Since the statistical analysis of $\mathbf{V}^{(k)}$ in terms of variance components for phenotype k has been explicitly discussed in Chapter 3, we will concentrate on the investigation of the lower triangular part of the symmetric variance-covariance matrix $\mathbf{V}$: $\mathbf{C}^{(kk')}$ $(k > k'$ and $k, k' = 1, \ldots, J)$ in this chapter.

Given the subject type indicator function $\mathcal{T}(\cdot)$ mapping the subject index to subject type (see Section 3.1 for more details), we can use the covariance for different paired response variables to construct $\mathbf{C}^{(kk')}$. Concerning the simplest bivariate case of two observed trait variables $\mathrm{Y}_{i,j}^{(k)}$ and $\mathrm{Y}_{i,j}^{(k')}$ from the same subject i with type index j for two phenotypes k and k', the covariance between these two data variables gives

the within-subject cross-phenotype covariance:

$$\begin{aligned}\mathbb{C}\text{ov}\mathrm{Y}^{(kk')} &= \mathbb{C}\text{ov}\begin{pmatrix}\mathrm{Y}_{i,j}^{(k)}\\ \mathrm{Y}_{i,j}^{(k')}\end{pmatrix}\\ &= \begin{pmatrix}\sigma_k^2 & \sigma_k\sigma_{k'}\rho_{\mathrm{P}}^{(kk')}\\ \sigma_k\sigma_{k'}\rho_{\mathrm{P}}^{(kk')} & \sigma_k^2\end{pmatrix},\end{aligned} \tag{4.2}$$

where $\sigma_k^2 = A^{(k)} + C^{(k)} + E^{(k)}$ is the phenotypic variance for phenotype k with the accompanying variance components $A^{(k)}$, $C^{(k)}$ and $E^{(k)}$, as introduced in Section 3.1, and $\rho_{\mathrm{P}}^{(kk')}$ denotes the phenotypic correlation between the paired phenotypes k and k' measuring the overall genetic and environmental relationships between these two phenotypes (Falconer and Mackay, 1996).

Concerning different subjects as well as paired phenotypes, the between-subject cross-phenotype covariance matrices can be constructed. As the above within-subject cross-phenotype covariance can be partitioned into genetic and environmental components, the cross-phenotype covariance between a pair of subjects can be derived based on the extent of genetic similarity between the paired subjects. For MZ twin pair j ($j = 1, \ldots, \frac{1}{2}n_{\mathrm{MZ}}$), $\mathcal{T}(2j-1) = \mathcal{T}(2j) = \mathrm{MZ}$. Since MZ twins have identical genotypes, the between-subject covariance between phenotypes k and k' for MZ twins is derived as

$$\begin{aligned}\mathbb{C}\text{ov}\mathrm{MZ}^{(kk')} &= \mathbb{C}\text{ov}\begin{pmatrix}\mathrm{Y}_{2j-1,j}^{(k)}\\ \mathrm{Y}_{2j,j}^{(k')}\end{pmatrix} = \mathbb{C}\text{ov}\begin{pmatrix}\mathrm{Y}_{2j,j}^{(k)}\\ \mathrm{Y}_{2j-1,j}^{(k')}\end{pmatrix}\\ &= \begin{pmatrix}\sigma_k^2 & \sigma_k\sigma_{k'}h_k h_{k'}\rho_{\mathrm{G}}^{(kk')}\\ \sigma_k\sigma_{k'}h_k h_{k'}\rho_{\mathrm{G}}^{(kk')} & \sigma_{k'}^2\end{pmatrix},\end{aligned} \tag{4.3}$$

where h_k represents the square root of heritability for phenotype k, $\rho_{\mathrm{G}}^{(kk')}$ denotes the genetic correlation between the paired phenotypes k and k' (see details below in the coming section), and the cross-phenotype covariance between MZ twins is coming from Equation (4.6) below. The genetic correlation measures the genetic association between these two phenotypes by quantifying the genetic variation common to both phenotypes (Glahn et al., 2012; Falconer and Mackay, 1996). For DZ twin pair j ($j = \frac{1}{2}n_{\mathrm{MZ}}+1, \ldots, \frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})$), $\mathcal{T}(2j-1) = \mathcal{T}(2j) = \mathrm{DZ}$, and the corresponding between-subject cross-phenotype covariance between phenotypes k and k' for DZ

twins is

$$\begin{aligned}\mathbb{C}\text{ov}\text{DZ}^{(kk')} &= \mathbb{C}\text{ov}\begin{pmatrix}\text{Y}^{(k)}_{2j-1,j}\\ \text{Y}^{(k')}_{2j,j}\end{pmatrix} = \mathbb{C}\text{ov}\begin{pmatrix}\text{Y}^{(k)}_{2j,j}\\ \text{Y}^{(k')}_{2j-1,j}\end{pmatrix}\\ &= \begin{pmatrix}\sigma^2_k & \frac{1}{2}\sigma_k\sigma_{k'}h_kh_{k'}\rho^{(kk')}_{\text{G}}\\ \frac{1}{2}\sigma_k\sigma_{k'}h_kh_{k'}\rho^{(kk')}_{\text{G}} & \sigma^2_{k'}\end{pmatrix}\end{aligned} \tag{4.4}$$

owing to the averagely 50% commonly shared genes for DZ twins (Falconer and Mackay, 1996). For a pair of unrelated individuals i and i' ($i \neq i'$ and $i, i' = 1, \ldots, n$) with subject type index of j and j' respectively, their between-subject covariance between the paired phenotypes k and k' is expressed as

$$\begin{aligned}\mathbb{C}\text{ov}\text{UN}^{(kk')} &= \mathbb{C}\text{ov}\begin{pmatrix}\text{Y}^{(k)}_{i,j}\\ \text{Y}^{(k')}_{i',j'}\end{pmatrix} = \mathbb{C}\text{ov}\begin{pmatrix}\text{Y}^{(k)}_{i',j'}\\ \text{Y}^{(k')}_{i,j}\end{pmatrix}\\ &= \begin{pmatrix}\sigma^2_k & 0\\ 0 & \sigma^2_{k'}\end{pmatrix}\end{aligned} \tag{4.5}$$

because of the independence between the unrelated subjects, where the formation of the possible unrelated subject pairs is as described in Section 3.1.

Based on these cross-phenotype covariance matrices including the within-subject cross-phenotype covariance (4.2) for individuals and between-subject cross-phenotype covariances (4.3), (4.4) and (4.5) for paired subjects from the MZ, DZ and UN (the unrelated) groups, the n-by-n block-diagonal cross-phenotype covariance matrix $\mathbf{C}^{(kk')}$ has the structure of

$$\mathbf{C}^{(kk')} = \begin{pmatrix}\mathbf{Q}^{(kk')}_{\text{MZ}} & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}\\ \vdots & \ddots & & & & & \vdots\\ \vdots & & \mathbf{Q}^{(kk')}_{\text{MZ}} & & & & \vdots\\ \vdots & & & \mathbf{Q}^{(kk')}_{\text{DZ}} & & & \vdots\\ \vdots & & & & \ddots & & \vdots\\ \vdots & & & & & \mathbf{Q}^{(kk')}_{\text{DZ}} & \vdots\\ \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{Q}^{(kk')}_{\text{S}}\end{pmatrix},$$

where the diagonal blocks of $\mathbf{Q}^{(kk')}_{\text{MZ}}$, $\mathbf{Q}^{(kk')}_{\text{DZ}}$ and $\mathbf{Q}^{(kk')}_{\text{S}}$ for phenotypes k and k' are

specified with the expressions as follows:

$$\mathbf{Q}_{\mathrm{MZ}}^{(kk')} = \sigma_k \sigma_{k'} \begin{pmatrix} \rho_{\mathrm{P}}^{(kk')} & h_k h_{k'} \rho_{\mathrm{G}}^{(kk')} \\ h_k h_{k'} \rho_{\mathrm{G}}^{(kk')} & \rho_{\mathrm{P}}^{(kk')} \end{pmatrix},$$

$$\mathbf{Q}_{\mathrm{DZ}}^{(kk')} = \sigma_k \sigma_{k'} \begin{pmatrix} \rho_{\mathrm{P}}^{(kk')} & \frac{1}{2} h_k h_{k'} \rho_{\mathrm{G}}^{(kk')} \\ \frac{1}{2} h_k h_{k'} \rho_{\mathrm{G}}^{(kk')} & \rho_{\mathrm{P}}^{(kk')} \end{pmatrix},$$

$$\begin{aligned} \mathbf{Q}_{\mathrm{S}}^{(kk')} &= \sigma_k \sigma_{k'} \begin{pmatrix} \rho_{\mathrm{P}}^{(kk')} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_{\mathrm{P}}^{(kk')} \end{pmatrix} \\ &= \sigma_k \sigma_{k'} \rho_{\mathrm{P}}^{(kk')} \mathbf{I}, \end{aligned}$$

for MZ twins, DZ twins and singletons respectively, where $\mathbf{I}$ is the $n_{\mathrm{S}} \times n_{\mathrm{S}}$ identity matrix. For each pair of phenotypes, there are $\frac{1}{2} n_{\mathrm{MZ}}$ $\mathbf{Q}_{\mathrm{MZ}}^{(kk')}$ blocks for $\frac{1}{2} n_{\mathrm{MZ}}$ MZ twin pairs at the top of diagonal blocks, then $\frac{1}{2} n_{\mathrm{DZ}}$ $\mathbf{Q}_{\mathrm{DZ}}^{(kk')}$ blocks for $\frac{1}{2} n_{\mathrm{DZ}}$ DZ twin pairs, and then a $\mathbf{Q}_{\mathrm{S}}^{(kk')}$ block for all singletons in the end. As detailed in Section 3.4, heritability h_k^2 $(k = 1, \ldots, J)$ for phenotype k can be estimated and tested using univariate LR-SD and LRT respectively, whereas the estimation of other unknown parameters and the inference on genetic correlation $\rho_{\mathrm{G}}^{(kk')}$ $(k > k'$ and $k, k' = 1, \ldots, J)$ for the paired phenotypes k and k' still need to be investigated and are the main concerns of this chapter.

Phenotypic and Genetic Correlations

The phenotypic correlation $\rho_{\mathrm{P}}^{(kk')}$ measures the correlation coefficient between a pair of phenotypes k and k' . Similar to decomposing the phenotypic variance into genetic and environmental components, the phenotypic correlation can also be partitioned into the additive genetic $\rho_{\mathrm{G}}^{(kk')}$ and environmental, denoted as $\rho_{\mathrm{E}}^{(kk')}$, parts, where the genetic correlation is conceptualized as the correlation coefficient between the common genetic influences underneath the paired traits considered (Searle, 1961; McGuffin et al., 2004).

By analogy with the decomposition of phenotypic variance $\sigma_k^2 = A_k + C_k + E_k$ for phenotype k, the within-subject covariance between two phenotypes k and k' with observations $\mathrm{Y}_{i,j}^{(k)}$ and $\mathrm{Y}_{i,j}^{(k')}$ can also be partitioned into the genetic and environ-

mental components:

$$\mathbb{C}\mathrm{ov}(\mathrm{Y}_{i,j}^{(k)}, \mathrm{Y}_{i,j}^{(k')}) = \mathbb{C}\mathrm{ov}_{\mathrm{G}} + \mathbb{C}\mathrm{ov}_{\mathrm{E}},$$

which is equivalent to

$$\begin{aligned}\sigma_k \sigma_{k'} \rho_{\mathrm{P}}^{(kk')} &= \sqrt{A_k A_{k'}} \rho_{\mathrm{G}}^{(kk')} + \sqrt{(\sigma_k^2 - A_k)(\sigma_{k'}^2 - A_{k'})} \rho_{\mathrm{E}}^{(kk')} \\ &= \sigma_k \sigma_{k'} h_k h_{k'} \rho_{\mathrm{G}}^{(kk')} + \sigma_k \sigma_{k'} \sqrt{(1 - h_k^2)(1 - h_{k'}^2)} \rho_{\mathrm{E}}^{(kk')} \end{aligned} \tag{4.6}$$

if the covariances are expressed in terms of phenotypic, genetic and environmental correlations. Therefore, the decomposition of the within-subject cross-phenotype covariance leads to the function of phenotypic correlation, which is comprised of heritabilities, genetic and environmental correlations and can be written as

$$\rho_{\mathrm{P}}^{(kk')} = h_k h_{k'} \rho_{\mathrm{G}}^{(kk')} + \sqrt{(1 - h_k^2)(1 - h_{k'}^2)} \rho_{\mathrm{E}}^{(kk')}. \tag{4.7}$$

The formula of genetic correlation can also be derived and expressed as

$$\rho_{\mathrm{G}}^{(kk')} = \frac{\mathbb{C}\mathrm{ov}_{\mathrm{G}}}{\sqrt{A_k A_{k'}}}.$$

This genetic correlation for quantitative traits is important in quantitative genetics for understanding the evolutionary process and predicting the genetically related response of traits, which is independent of the heritability of the paired phenotypic traits. That is, no matter how heritable the two traits are, the genetic correlation could be anywhere from -1 to 1; a high genetic correlation indicates that the genes found to have genetic influence on one trait are very likely to be associated with another, while a zero genetic correlation implies the genetic influence on one trait is not associated with that on another trait (McGuffin et al., 2004). However, when either trait heritability is exactly zero, the genetic correlation should be zero as well since there are no genetic effects.

Endophenotype Ranking Value

To study the neuropsychiatric diseases, particularly to understand the complex diseases, the key thing is finding the appropriate phenotypes. The endophenotypes were proposed to act as the intermediate phenotypes linking between illness and genotype (Goldstein and Klein, 2014). An endophenotype is conceptualized as the internal heritable characteristic that is measurable but not always observable, and must be genetically associated with the illness, however, it is difficult to express it precisely in the mathematical form. The optimal endophenotypes are less hetero-

geneous and can also be found in unaffected relatives. Although the endophenotypes are defined to be state-independent, people use them to predict the clinical states and outcomes (Gottesman and Gould, 2003; Gotlib and Hamilton, 2012).

The endophenotype ranking value (ERV) is an index measuring the potential genetic utility of the endophenotype for a given illness, which varies between 0 and 1 with higher value representing stronger genetic influence. The ERV can be used to efficiently assess and exhaustively rank a large number of potential neuroimaging endophenotypes, and is easily applicable to any heritable diseases and relevant traits. The identification of possible endophenotypes can be done by assessing the ERV before conducting the molecular genetic analysis (Glahn et al., 2012; Gotlib and Hamilton, 2012).

The ERV for a given pair of phenotypes k and k' is defined as the absolute value of the product of the square root of two heritabilities (h_k and $h_{k'}$) and the genetic correlation $\left(\rho_{\mathrm{G}}^{(kk')}\right)$ for the paired phenotypes k and k', which can be expressed in a formula:

$$\mathrm{ERV}^{(kk')} = |h_k h_{k'} \rho_{\mathrm{G}}^{(kk')}| \tag{4.8}$$

(Glahn et al., 2012). The value of ERV increases while either heritability of these two phenotypes raises or the genetic correlation between this pair of phenotypes enlarges. The definition of the ERV quantity balances the genetic strengths for the paired phenotypes by importing the measure of their genetic relationship $\rho_{\mathrm{G}}^{(kk')}$.

4.2 Accelerated Multivariate Heritability Analysis

The idea of constructing linear regression model with SD's, estimating unknown variance components A, C and E using NNLS algorithm in the univariate heritability analysis for a single observable trait (see details in Section 3.4) can be analogously applied to the multivariate modeling of multiple phenotypes altogether. This section contributes to the model construction and parameter inference for the bivariate case and the generalization from the bivariate modeling to multivariate modeling is fairly straightforward.

4.2.1 Bivariate Linear Regression with Squared Differences

In this section, the univariate LR-SD method is extended to be used for bivariate inference with a pair of phenotypes. Compared with the univariate modeling, the

bivariate case can similarly build the linear regression model using the bivariate SD's that can be calculated from different phenotypes within individuals or between subjects. To take the subject-specific variability into consideration, the between-subject covariates are measured and included in the multivariate GLM (4.1).

Linear Regression Model

As stated above in Section 3.4.1, for each phenotype, the differences among subjects arisen with respect to the included covariates may lead to the unwanted between-subject variation that needs to be controlled. The within-phenotype differences explained by the covariates or cross-phenotype mean variation should be excluded prior to the later analysis (if possible) so that the genetic effects will not be confounded with these redundant nuisance effects during the statistical analysis.

Moreover, the variances of unobservable subject-specific errors for distinct phenotypes may differ substantially, and the use of the phenotypic datasets with different variation may result in confounding and possibly lead to misleading conclusions. This problem is described as variance instability, and the purpose of normalization is to remove the non-negligible variation of the phenotypic variances and resolve the instability of variance. For each phenotype, we attempt to further variance-normalize the derived demeaned data (or residual) by dividing the residual by its sample standard deviation (a scalar quantity) to account for different variances in the residuals obtained from multiple phenotypes. This variance normalization deals with the probable variation in phenotypic variances and allows those residuals to be compared in the same scale. The variance-normalized (or scaled) residual is a quotient, which is written as

$$\widetilde{\mathbf{e}}^{(k)} = \frac{\mathbf{e}^{(k)}}{\mathrm{s}\Big(\mathbf{e}^{(k)}\Big)} \tag{4.9}$$

for phenotype k ($k = 1, \ldots, J$) from all subjects, where the residual using OLS is expressed as

$$\mathbf{e}^{(k)} = \mathbf{Y}^{(k)} - \mathbf{X}^{(k)}(\mathbf{X}^{(k)\mathrm{T}}\mathbf{X}^{(k)})^{-1}\mathbf{X}^{(k)\mathrm{T}}\mathbf{Y}^{(k)},$$

and $\mathrm{s}\Big(\mathbf{e}^{(k)}\Big)$ is the sample standard deviation of $\mathbf{e}^{(k)}$. The key reason of applying variance normalization is to dispose the disagreement between phenotypes in terms of their variances. Even though it is suggested to variance-normalize the residual $\mathbf{e}^{(k)}$ in order to obtain unit variance, the step of variance normalization can be omitted if the variances of phenotypic data appear to be indistinguishable and the potential variation in phenotypic variances seems non-existent, and thus phenotypic

variance σ_k^2 will be kept for generality in the later derivation.

Consider the bivariate modeling of $J = 2$ phenotypes. We denote the associated data vectors from two phenotypes as $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$, and then the multivariate GLM model (4.1) is simply expressed as

$$\begin{pmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}^{(2)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}^{(1)} \\ \boldsymbol{\epsilon}^{(2)} \end{pmatrix}, \ \begin{pmatrix} \boldsymbol{\epsilon}^{(1)} \\ \boldsymbol{\epsilon}^{(2)} \end{pmatrix} \sim \mathbb{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V}^{(1)} & \mathbf{C}^{(12)} \\ \mathbf{C}^{(21)} & \mathbf{V}^{(2)} \end{pmatrix} \right).$$

Prior to the construction of linear regression model using bivariate SD's for inference, each phenotypic data vector $\mathbf{Y}^{(k)}$ $(k = 1, 2)$ is mean-centered (covariate-adjusted) and variance-normalized by computing the scaled residual with Equation (4.9), and the resulting scaled residuals will be treated as the observed data for the later linear regression model construction with bivariate SD's after normalization. For simplicity of notation going forward, we write $\mathbf{Y}^{(k)}$ instead of $\widetilde{\mathbf{e}}^{(k)}$ for the normalized data. This represents an assumption that each entry of the phenotypic (or element-wise phenotypic) data vector $\mathbf{Y}^{(k)}$ has zero mean and unit variance after data transformation using mean-centering and variance normalization; we neglect any correlation induced by using these data normalization methods.

To analyze the phenotypic correlation, we compute the within-subject bivariate SD's using the data points of the two phenotypes from a single subject and employ the within-subject cross-phenotype covariance matrix (4.2) to calculate the expected value of each of these within-subject SD's:

$$\begin{aligned} \mathbb{E}\left[(\mathrm{Y}_{i,j}^{(1)} - \mathrm{Y}_{i,j}^{(2)})^2\right] &= \mathbb{V}\mathrm{ar}\left(\mathrm{Y}_{i,j}^{(1)} - \mathrm{Y}_{i,j}^{(2)}\right) \\ &= \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{\mathrm{P}}^{(12)} \end{aligned} \tag{4.10}$$

for the subject i with type index j.

For the analysis of genetic correlation, we consider computing each between-subject bivariate SD using the data from different subjects and across paired phenotypes, where the between-subject SD's can be further separated into 3 groups of MZ, DZ and UN (unrelated subject pairs). The between-subject cross-phenotype covariance matrices for MZ twin pairs (4.3), DZ twin pairs (4.4) and the unrelated subject pairs (4.5) are utilized with the use of the basic properties of variance operator to

compute the expectations of these 3 groups of between-subject SD's; those are

$$\begin{aligned}\mathbb{E}\left[(\mathrm{Y}^{(1)}_{2j-1,j}-\mathrm{Y}^{(2)}_{2j,j})^2\right] &= \mathbb{V}\mathrm{ar}\left(\mathrm{Y}^{(1)}_{2j-1,j}-\mathrm{Y}^{(2)}_{2j,j}\right)\\ &= \sigma_1^2+\sigma_2^2-2\sigma_1\sigma_2 h_1 h_2 \rho_{\mathrm{G}}^{(12)} \end{aligned} \tag{4.11}$$

for MZ twin pair j ($j=1,\ldots,\frac{1}{2}n_{\mathrm{MZ}}$),

$$\begin{aligned}\mathbb{E}\left[(\mathrm{Y}^{(1)}_{2j-1,j}-\mathrm{Y}^{(2)}_{2j,j})^2\right] &= \mathbb{V}\mathrm{ar}\left(\mathrm{Y}^{(1)}_{2j-1,j}-\mathrm{Y}^{(2)}_{2j,j}\right)\\ &= \sigma_1^2+\sigma_2^2-\sigma_1\sigma_2 h_1 h_2 \rho_{\mathrm{G}}^{(12)} \end{aligned} \tag{4.12}$$

for DZ twin pair j ($j=\frac{1}{2}n_{\mathrm{MZ}}+1,\ldots,\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})$), and for the remaining unrelated subject pair of $\mathrm{Y}_{i,j}$ and $\mathrm{Y}_{i',j'}$,

$$\begin{aligned}\mathbb{E}\left[(\mathrm{Y}^{(1)}_{i,j}-\mathrm{Y}^{(2)}_{i',j'})^2\right] &= \mathbb{V}\mathrm{ar}\left(\mathrm{Y}^{(1)}_{i,j}-\mathrm{Y}^{(2)}_{i',j'}\right)\\ &= \sigma_1^2+\sigma_2^2 \end{aligned} \tag{4.13}$$

for the paired subjects i and i' with type indices of j and j' respectively.

Equations (4.10), (4.11), (4.12) and (4.13) describe the expected values for all n^2 bivariate SD's between the paired phenotypes within or between subjects. As we are more concerned about the phenotypic and genetic correlations, we further attempt to take the difference between (4.10), (4.11), (4.12) and (4.13) by subtraction so as to exclude the nuisance term of $\sigma_1^2+\sigma_2^2$ and hopefully improve the estimation accuracy. For the estimation of phenotypic correlation $\rho_{\mathrm{P}}^{(12)}$, the mean difference between cross-phenotype SD's for the UN group and within-subject cross-phenotype SD's, i.e., (4.13) − (4.10), is calculated and written as

$$\begin{aligned}\mathbb{E}\left[(\mathrm{Y}^{(1)}_{i',j'}-\mathrm{Y}^{(2)}_{i'',j''})^2-(\mathrm{Y}^{(1)}_{i,j}-\mathrm{Y}^{(2)}_{i,j})^2\right] &= \mathbb{E}\left[(\mathrm{Y}^{(1)}_{i',j'}-\mathrm{Y}^{(2)}_{i'',j''})^2\right]-\mathbb{E}\left[(\mathrm{Y}^{(1)}_{i,j}-\mathrm{Y}^{(2)}_{i,j})^2\right]\\ &= 2\sigma_1\sigma_2\rho_{\mathrm{P}}^{(12)}, \end{aligned} \tag{4.14}$$

where i' and i'' represent a pair of unrelated subjects with type indices of j' and j'' from the UN group, and the subject i with index j denotes an arbitrary individual within the sample. While for the analysis of genetic correlation $\rho_{\mathrm{G}}^{(12)}$, the mean differences between cross-phenotype SD's for the UN group and those for twin pairs of MZ and DZ, i.e., (4.13) − (4.11) and (4.13) − (4.12) for the MZ and DZ groups

respectively, are derived with the expression of

$$\begin{aligned}
&\mathbb{E}\left[(\mathrm{Y}^{(1)}_{i',j'} - \mathrm{Y}^{(2)}_{i'',j''})^2 - (\mathrm{Y}^{(1)}_{2\bar{j}-1,\bar{j}} - \mathrm{Y}^{(2)}_{2\bar{j},\bar{j}})^2\right] \\
&= \mathbb{E}\left[(\mathrm{Y}^{(1)}_{i',j'} - \mathrm{Y}^{(2)}_{i'',j''})^2\right] - \mathbb{E}\left[(\mathrm{Y}^{(1)}_{2\bar{j}-1,\bar{j}} - \mathrm{Y}^{(2)}_{2\bar{j},\bar{j}})^2\right] \\
&= \begin{cases} 2\sigma_1\sigma_2 h_1 h_2 \rho^{(12)}_{\mathrm{G}} & \text{for MZ twins} \\ \sigma_1\sigma_2 h_1 h_2 \rho^{(12)}_{\mathrm{G}} & \text{for DZ twins} \end{cases}
\end{aligned} \tag{4.15}$$

where i' and i'' represent a pair of unrelated subjects with type indices of j' and j'', and $2\bar{j}-1$ and $2\bar{j}$ are twins from the pair $\bar{j}$ regardless of MZ or DZ.

Therefore we can specify the mean structure of two linear regression models with bivariate SD's for the inference of correlations, one for phenotypic correlation using Equation (4.14) with $n(n^2 - n - n_{\mathrm{MZ}} - n_{\mathrm{DZ}})$ observations:

$$\mathbb{E}\begin{pmatrix} (\mathrm{Y}^{(1)}_{1,1} - \mathrm{Y}^{(2)}_{3,2})^2 - (\mathrm{Y}^{(1)}_{1,1} - \mathrm{Y}^{(2)}_{1,1})^2 \\ \vdots \\ (\mathrm{Y}^{(1)}_{n-1,0} - \mathrm{Y}^{(2)}_{n,0})^2 - (\mathrm{Y}^{(1)}_{n,0} - \mathrm{Y}^{(2)}_{n,0})^2 \end{pmatrix} = \begin{pmatrix} 2 \\ \vdots \\ 2 \end{pmatrix} \left(\sigma_1\sigma_2\rho^{(12)}_{\mathrm{P}}\right),$$

and another for genetic correlation using Equations (4.15) with totally $(n_{\mathrm{MZ}} + n_{\mathrm{DZ}})(n^2 - n - n_{\mathrm{MZ}} - n_{\mathrm{DZ}})$ observations:

$$\mathbb{E}\begin{pmatrix} (\mathrm{Y}^{(1)}_{1,1} - \mathrm{Y}^{(2)}_{3,2})^2 - (\mathrm{Y}^{(1)}_{1,1} - \mathrm{Y}^{(2)}_{2,1})^2 \\ \vdots \\ (\mathrm{Y}^{(1)}_{n-1,0} - \mathrm{Y}^{(2)}_{n,0})^2 - (\mathrm{Y}^{(1)}_{n_{\mathrm{MZ}}-1,\frac{1}{2}n_{\mathrm{MZ}}} - \mathrm{Y}^{(2)}_{n_{\mathrm{MZ}},\frac{1}{2}n_{\mathrm{MZ}}})^2 \\ (\mathrm{Y}^{(1)}_{1,1} - \mathrm{Y}^{(2)}_{3,2})^2 - (\mathrm{Y}^{(1)}_{n_{\mathrm{MZ}}+1,\frac{1}{2}n_{\mathrm{MZ}}+1} - \mathrm{Y}^{(2)}_{n_{\mathrm{MZ}}+2,\frac{1}{2}n_{\mathrm{MZ}}+1})^2 \\ \vdots \\ (\mathrm{Y}^{(1)}_{n-1,0} - \mathrm{Y}^{(2)}_{n,0})^2 - (\mathrm{Y}^{(1)}_{n_{\mathrm{MZ}}+n_{\mathrm{DZ}}-1,\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})} - \mathrm{Y}^{(2)}_{n_{\mathrm{MZ}}+n_{\mathrm{DZ}},\frac{1}{2}(n_{\mathrm{MZ}}+n_{\mathrm{DZ}})})^2 \end{pmatrix} =$$
$$\begin{pmatrix} 2 \\ \vdots \\ 2 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \left(\sigma_1\sigma_2 h_1 h_2 \rho^{(12)}_{\mathrm{G}}\right).$$

For simplicity, these two linear regression models with cross-phenotype bivariate SD's can be simply denoted as

$$\mathbb{E}\left[\mathbf{D}_{\mathrm{P}}^{(12)}\right] = \mathbf{Z}_{\mathrm{P}}^{(12)}\eta_{\mathrm{P}}^{(12)},$$

$$\mathbb{E}\left[\mathbf{D}_{\mathrm{G}}^{(12)}\right] = \mathbf{Z}_{\mathrm{G}}^{(12)}\eta_{\mathrm{G}}^{(12)},$$

where $\eta_{\mathrm{P}}^{(12)} = \sigma_1\sigma_2\rho_{\mathrm{P}}^{(12)}$ and $\eta_{\mathrm{G}}^{(12)} = \sigma_1\sigma_2 h_1 h_2 \rho_{\mathrm{G}}^{(12)}$.

4.2.2 Parameter Estimation

The estimation of the unknown parameters $\eta_{\mathrm{P}}^{(12)}$ and $\eta_{\mathrm{G}}^{(12)}$ is made with the simple OLS approach. The resulting OLS estimates $\widehat{\eta}_{\mathrm{P}}^{(12)}$ and $\widehat{\eta}_{\mathrm{G}}^{(12)}$ can be further divided by the terms of $\widehat{\sigma}_1\widehat{\sigma}_2$ and $\widehat{\sigma}_1\widehat{\sigma}_2\widehat{h}_1\widehat{h}_2$ (when $\widehat{h}_k^2 \neq 0$ $(k = 1, 2)$) separately to yield the bivariate LR-SD estimates of phenotypic and genetic correlations, expressed as

$$\widehat{\rho}_{\mathrm{P}}^{(12)} = \frac{\widehat{\eta}_{\mathrm{P}}^{(12)}}{\widehat{\sigma}_1\widehat{\sigma}_2}, \tag{4.16}$$

$$\widehat{\rho}_{\mathrm{G}}^{(12)} = \frac{\widehat{\eta}_{\mathrm{G}}^{(12)}}{\widehat{\sigma}_1\widehat{\sigma}_2\widehat{h}_1\widehat{h}_2}, \tag{4.17}$$

where $\widehat{\sigma}_k$ and $\widehat{h}_k$ $(k = 1, 2)$ are the estimates of phenotypic standard deviation and the square root of heritability obtained from the univariate LR-SD method.

4.2.3 Hypothesis Testing

The heritability measuring the genetic sources affecting a phenotype is generally of more concern, and the genetic correlation describing the commonly shared genetic causes between the paired phenotypes is also the parameter of interest. With regard to Formula (4.8) for the computation of ERV, the ERV quantity has a non-zero value when both genetic correlation and heritabilities are non-zero and vice versa. For each pair of phenotypes, testing the null hypothesis $\mathrm{H}_0 : \rho_{\mathrm{G}}^{(12)} = 0,\ h_1^2 = 0,$ or $h_2^2 = 0$ is the equivalent of assessing the statement of $\mathrm{ERV}^{(12)} = 0$, which gives the null hypothesis $\mathrm{H}_0 : \mathrm{ERV}^{(12)} = 0$ that needs to be examined during the significance testing.

Regarding the selection of test statistic, we firstly consider using the LRT statistic by comparing the fit of the null model and that of the alternative model, as stated in Section 3.4.3, with the expression of

$$\mathrm{T}_1^{(12)} = -2 \times \left[\ell(\widehat{\mathrm{ERV}}_0^{(12)}|\mathbf{Y}) - \ell(\widehat{\mathrm{ERV}}_1^{(12)}|\mathbf{Y})\right],$$

where $\widehat{\text{ERV}}_0^{(12)}$ and $\widehat{\text{ERV}}_1^{(12)}$ are estimated from the null and alternative models respectively. However, this commonly used LRT statistic is computationally very intensive for moderate to large sample sizes and nearly infeasible for the use of permutation inference. Therefore, we consider to employ the ERV estimator:

$$\text{T}_2^{(12)} = \widehat{\text{ERV}}^{(12)},$$

as an alternative test statistic. The validity of this test statistic will be investigated with simulations in the later section.

4.2.4 Permutation Framework

The exact sampling distribution of the chosen test statistic (either the LRT statistic or ERV estimator) under the null hypothesis $\text{H}_0 : \text{ERV}^{(12)} = 0$ is unknown. The permutation test provides a simple way to estimate the exact null distribution of these test statistics. When the null hypothesis is true, the between-subject cross-phenotype covariance matrices for MZ and DZ twins are equivalent, and thus exchangeable. By randomly shuffling the labels of MZ and DZ for those between-subject cross-phenotype covariance matrices in the multivariate GLM model (4.1), we generate a new variance-covariance matrix $\mathbf{V}$ within the permuted GLM model that should look like the original one, assuming the null hypothesis is true. The final step of our permutation scheme is to construct the empirical null distribution and the ranking of the original value among all shuffled values of the test statistic gives the permutation-based p-value that can be used to interpret the test statistic and quantify the significance of the test.

4.2.5 Simulation-based Analysis

This section contributes to the simulation-based evaluation of the bivariate LR-SD method in terms of estimation accuracy with different simulation settings. The newly proposed test statistic of the ERV is also examined with simulations.

Simulation Setting

We start the Monte Carlo simulation evaluations by setting up the simulation settings. Consider there are a pair of phenotypes with distinct phenotypic variances of $\sigma_1^2 = 1$ and $\sigma_2^2 = 100$ and with different phenotypic averages generated from a uniform distribution on the interval $[0, 100]$, and assume that these two phenotypic traits have identical magnitude of heritability. The selected settings of heritability

and normalized variance components $(A, C, E)^{\mathrm{T}}$ are shown in Table 4.1. The above 1D simulation results show that with the same amount of unique environmental factor E the existence of common environmental factor C lowers the estimation accuracy and decreases the statistical power, so we consider the $(A, C, E)^{\mathrm{T}}$ settings with the strongest effect size, i.e., the largest heritability, by setting $C = 0$ for each parameter value of E. As mentioned earlier, the E effect is always existent and often large in reality, hence three settings of E including $E = \frac{1}{2}, \frac{3}{4}, 1$ are chosen. In Table 4.2, the genetic and phenotypic correlations are shown in pairs. The correlation coefficient ranges from -1 to 1, so we set genetic correlation to be $\rho_{\mathrm{G}} = -1, -0.5, 0, 0.5, 1$ to account for its variability. Because of Equation (4.7) and the constraint on correlations (between -1 and 1), we assign $\rho_{\mathrm{P}} = -0.5, 0, 0.5$ to phenotypic correlation. There are entirely 3 samples considered of size $n = 100, 200, 400$, where each sample includes twin data only, and MZ and DZ twin pairs are the same in number (e.g., the sample of size $n = 100$ contains 25 MZ and 25 DZ twin pairs). In total, $n_{\mathrm{Rlz}} = 1000$ realizations were performed, and N = 1000 permutations were implemented for each realization.

Table 4.1: 3 parameter settings of heritability h^2 and normalized variance components $(A, C, E)^{\mathrm{T}}$.

h^2	A	C	E
0	0	0	1
1/4	1/4	0	3/4
1/2	1/2	0	1/2

Table 4.2: The entire 13 parameter settings of genetic and phenotypic correlations in pairs of $(\rho_{\mathrm{G}}, \rho_{\mathrm{P}})$.

ρ_{G}	-1	-1	-0.5	-0.5	-0.5	0	0	0	0.5	0.5	0.5	1	1
ρ_{P}	-0.5	0	-0.5	0	0.5	-0.5	0	0.5	-0.5	0	0.5	0	0.5

Evaluation Results

In this section, the simulation-based evaluation results for bivariate LR-SD are shown in terms of estimation accuracy and statistical validity.

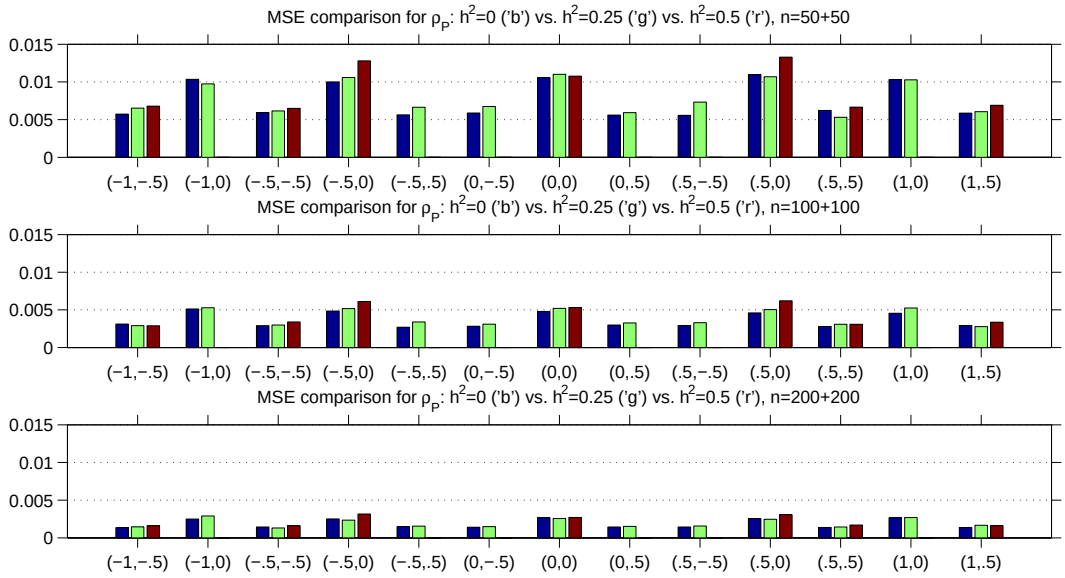


Figure 4.1: The MSE comparison of bivariate LR-SD for the estimation of phenotypic correlation ρ_{P} for different heritability settings (shown in Table 4.1): $h^2 = 0$ ('b': blue), $h^2 = 0.25$ ('g': green) and $h^2 = 0.5$ ('r': red). Comma ordered pairs on x-axis correspond to the paired values of genetic correlation ρ_{G} and ρ_{P}, i.e., $(\rho_{\mathrm{G}}, \rho_{\mathrm{P}})$; see Table 4.2 for exact parameter settings used.

Estimation Accuracy Figure 4.1 illustrates the MSE comparison of the estimators for phenotypic correlation ρ_{P} using bivariate LR-SD for different parameter settings. Compared with the true ρ_{P} magnitudes, the measures of MSE for these ρ_{P} estimator are negligible. In Figure 4.2, different simulation settings are compared in terms of the MSE values for the estimator for genetic correlation ρ_{G} obtained using bivariate LR-SD. It is shown that the realizations generated with comparatively smaller heritability values have relatively larger MSE values, which implies that the resulting ρ_{G} estimator is more biased or more variable for less heritable phenotypes. The comparison of Figures 4.1 and 4.2 indicates that bivariate LR-SD can derive more accurate ρ_{P} estimates than ρ_{G} estimates, which can be seen from Equations (4.16) and (4.17) computing the estimators for ρ_{P} and ρ_{G} that the additional estimation for two heritabilities using bivariate LR-SD can bring more bias or variation and result in less correct estimator for ρ_{G}. As expected, Figures 4.1 and 4.2 both depict that a larger sample size induces better estimation accuracy with smaller MSE value.

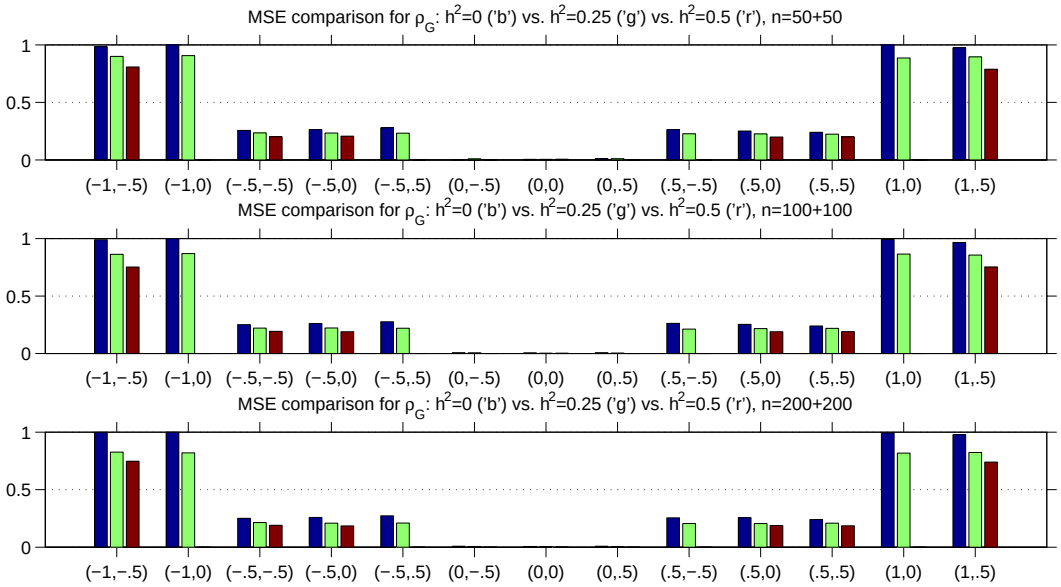


Figure 4.2: The MSE comparison of bivariate LR-SD for the estimation of genetic correlation ρ_{G} for different heritability settings (shown in Table 4.1): $h^2 = 0$ ('b': blue), $h^2 = 0.25$ ('g': green) and $h^2 = 0.5$ ('r': red). Comma ordered pairs on x-axis correspond to the paired values of ρ_{G} and phenotypic correlation ρ_{P}, i.e., $(\rho_{\mathrm{G}}, \rho_{\mathrm{P}})$; see Table 4.2 for exact parameter settings used.

Statistical Validity Figure 4.3 shows the permutation-based rejection rate, expressed as a percentage, for testing the null hypothesis $H_0 : ERV = 0$ using the test statistic of the ERV estimator for different parameter settings shown in Tables 4.1 and 4.2. The red dash-dotted lines in Figure 4.3 represent the lower and upper bounds of the 95% binomial proportion confidence interval. The null hypothesis is true when ERV = 0, or equivalently, when $h^2 = 0$ or $\rho_G = 0$. In Figure 4.3, the simulation settings with $h^2 = 0$ are shown in the blue bars, and the settings with $\rho_G = 0$ are shown in the middle with x-axis markers $(0, *)$. For nearly all simulation settings, the rate of rejecting the null hypothesis when it is actually true, i.e., the estimated FPR, lies within or below the 95% binomial proportion confidence interval, which reveals that the test is valid with the use of the ERV statistic. Other simulation results also indicate that the ERV estimator has small bias and variance with negligible MSE magnitude for all simulation settings, which encourages the use of the ERV as the test statistic.

4.2.6 Real Data Analysis

Aside from fMRI data, the univariate and bivariate LR-SD approaches can also be applied to any other type of neuroimaging data. In this section, we apply the bivariate LR-SD method to the diffusion tensor imaging (DTI) data for the genetic analysis. With the use of a recently developed diffusion MRI technology, DTI, based on the diffusion anisotropy of water molecules, is the most commonly used non-invasive technique studying the cerebral white matter structure and is sensitive to fiber tract integrity and white matter microstructure. The sensitive and relatively reliable measure of fractional anisotropy (FA), derived from DTI images, is often used to describe the anisotropic degree of a diffusion process.

Data Acquisition

The Human Connectome Project (HCP), with the publicly and freely available datasets supplied for the studies of brain structure, function and connectivity, provides data from various imaging modalities including structural MRI, fMRI and diffusion MRI. The recent HCP 500 Subjects Release [1] includes the 3T MR imaging data from 523 healthy adults. Prior to the genetic inference, we extract the data from twins out of these 523 subjects to form a sample of $n = 178$ subjects including $n_{MZ} = 98$ MZ twins (49 pairs) and $n_{DZ} = 80$ DZ twins (40 pairs) from 89 families. Among these 178 subjects, there are 126 females and 52 males with an age range of 22–36

[1] `http://www.humanconnectome.org/documentation/S500/`

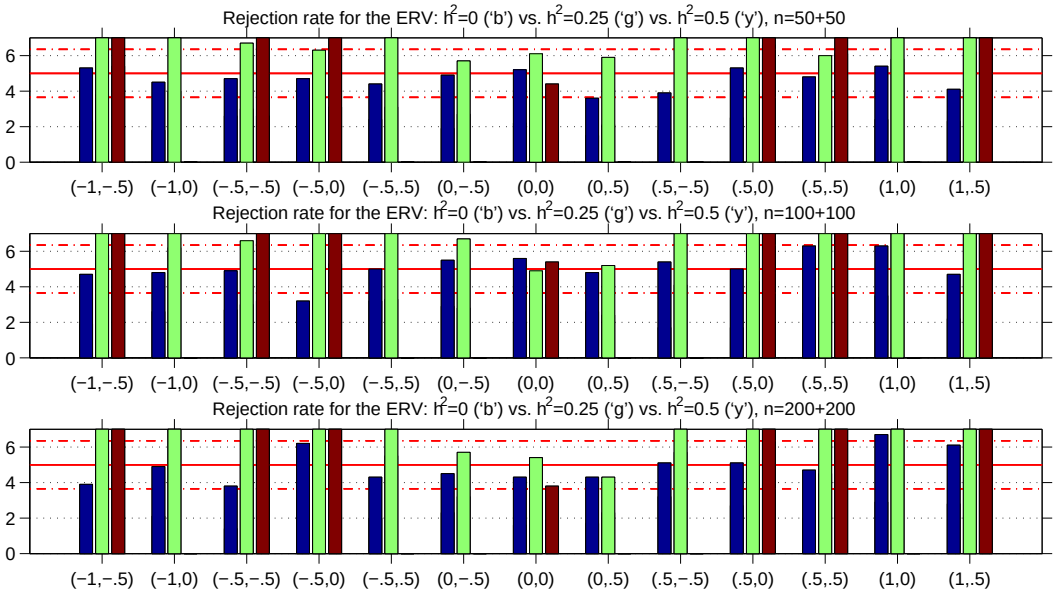


Figure 4.3: The comparison of the rejection rate (in percent) for the null hypothesis $\mathrm{H}_0 : \mathrm{ERV} = 0$ at level $\alpha = 0.05$ using the ERV as the test statistic for different heritability extents (shown in Table 4.1): $h^2 = 0$ ('b': blue), $h^2 = 0.25$ ('g': green) and $h^2 = 0.5$ ('r': red). Comma ordered pairs on x-axis correspond to the paired values of ρ_{G} and ρ_{P}, i.e., $(\rho_{\mathrm{G}}, \rho_{\mathrm{P}})$; see Table 4.2 for exact parameter settings used. The red dash-dotted lines show the lower and upper bounds of the 95% binomial proportion confidence interval. The FPR for the null settings should be 0.05, but its estimates can vary within the 95% binomial proportion confidence interval $[0.0365, 0.0635]$ for $n_{\mathrm{Rlz}} = 1000$ simulations.

(mean±SD : 29.8±2.9). With the introduction of the DTI image analysis process by Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) [2], the raw DTI scans were pre-processed and quality-controlled to achieve the regional data for each subject. The FA images of 400 randomly selected healthy adults from 4 independently collected datasets, 100 from each, were used to create a common template (Jahanshad et al., 2013). These FA images were aligned to the Johns Hopkins University (JHU) DTI white matter atlas (Mori et al., 2008) using FSL [3]. A target image was created from these aligned images and skeletonized with tract-based spatial statistics (TBSS) analysis (Smith et al., 2006), and then each HCP FA image from the above-mentioned sample of $n = 178$ subjects can be projected onto the template skeleton (Jahanshad et al., 2013). For each subject, totally 36 ROI's were parcellated from the ENIGMA template in ICBM space based on multi-subject JHU white matter parcellation atlas (Mori et al., 2008), and mean FA for these ROI's was extracted as the regional phenotypic measure (Jahanshad et al., 2013). Both univariate and bivariate LR-SD methods are employed for the estimation of genetic factors including heritability and genetic correlation, the ERV test statistic is adopted within the permutation framework to test the combined significance of the genetic influences, and the FWE correction and FDR control of p-values are applied to generate the FWE-corrected and FDR-controlled p-value images describing the brain reginal connectivity.

Results

With the use of reginal mean FA values from those 36 ROI's as the data from different phenotypic traits, we find non-zero heritability measures for all ROI's with a range of 0.04–0.85. Figures 4.4 and 4.5 illustrate the brain connectivity matrices based on the FWE-corrected and FDR-controlled p-values for all ROI pairs respectively. Among these 36 ROI's, there are totally $36 \times 35 = 1260$ pairs of ROI's (630 distinct) taken into consideration. The most significant FWE-corrected p-value is 0.001 with 3 significant ROI pairs, and the best attainable FDR is 0.014 with 199 ROI pairs found to be significant, which reveals that the FDR control possesses more statistical power while the FWE correction detects more significant results.

[2] `http://enigma.ini.usc.edu/protocols/dti-protocols/`
[3] `http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/`

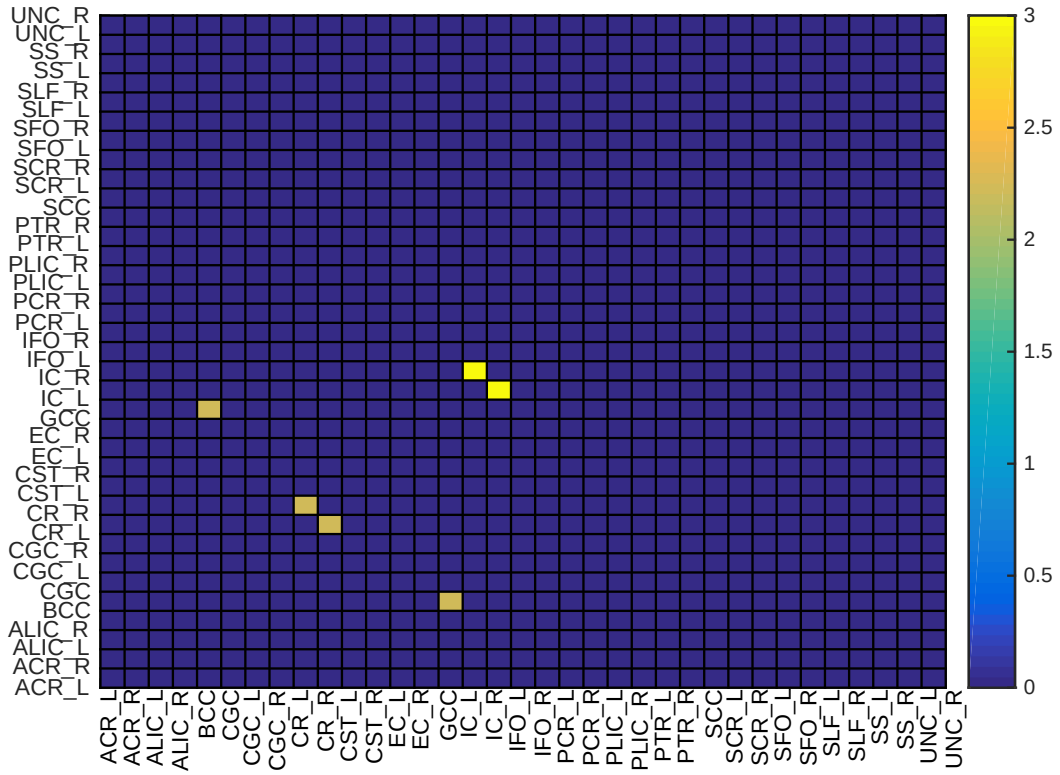


Figure 4.4: The brain connectivity matrix in terms of the log transformed FWE-corrected p-values (i.e., $-\log_{10}(p_{\text{FWE}})$) derived from 36 ROI's and totally $36 \times 35 = 1260$ ROI pairs using their mean FA values as the regional measures. Each marker on x and y axes corresponds to a ROI.

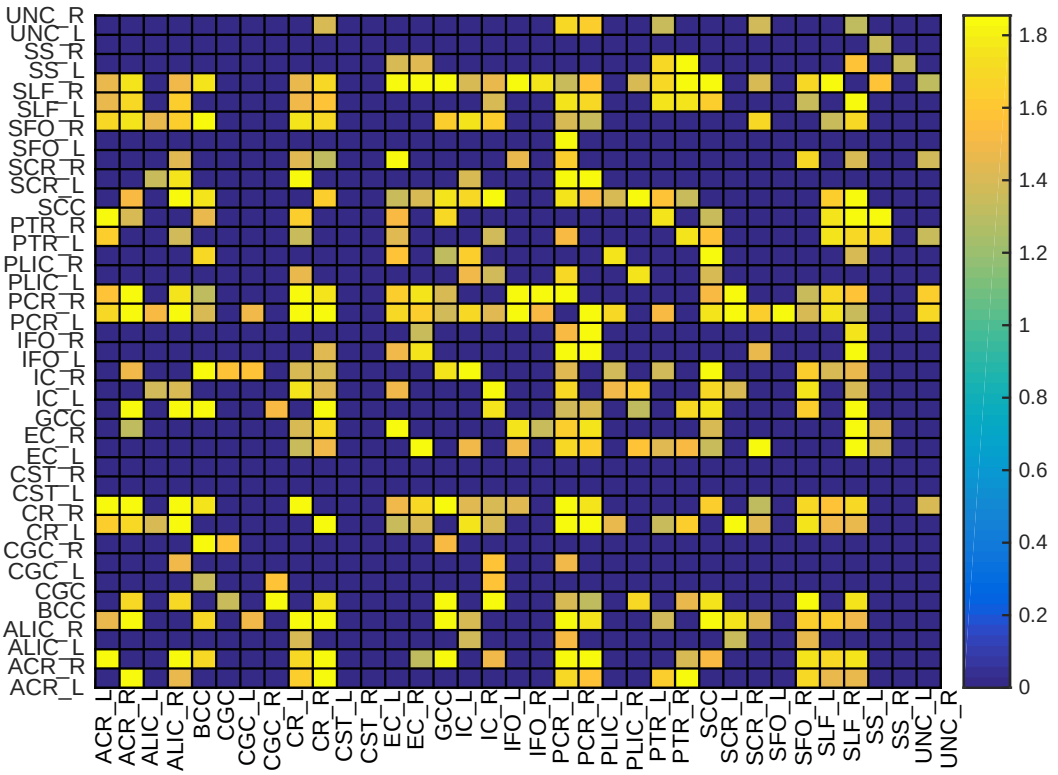


Figure 4.5: The brain connectivity matrix in terms of the log transformed FDR-controlled p-values (i.e., $-\log_{10}(p_{\text{FDR}})$) derived from 36 ROI's and totally $36 \times 35 = 1260$ ROI pairs using their mean FA values as the regional measures. Each marker on x and y axes corresponds to a ROI.

4.3 Fast Multivariate Heritability Analysis

Heritability, the proportion of variability attributable to genetic sources, is a vital quantitative genetic measure and, in particular, non-zero heritability is needed to certify a trait as a "phenotype". However heritability can also be used as a general measure of biological validity, e.g., ranking different pre-processing techniques by heritability of the resulting phenotypes. While such comparisons can be done element-wise over a high-dimensional phenotype (e.g., by voxels or surface elements), a whole-phenotype summary of element-wise multivariate heritability can simplify the comparisons.

In this section, we investigate a simple measure of aggregate heritability that is easy and extremely fast to compute and involves no ACE model fitting. We derive analytical results that show this aggregate measure is closely related to the average of element-wise heritability. In addition to validating our analytical results with simulations, we illustrate this method on 22 different phenotypes based on the data of 196 subjects from the publicly released Human Connectome Project (HCP) (Van Essen et al., 2013), comparing the ranking of this fast aggregate method to the slower traditional ACE-based estimates of average heritability.

4.3.1 Pair-wise Correlation and Aggregate Heritability

Assume that there are J elements in total in an fMRI data image from an imaging twin study. For a specified high-dimensional phenotype, each data element of this phenotype is treated as a single element-wise observable trait associated with the univariate heritability inference, and thus we have entirely J element-wise traits of this high-dimensional phenotype to be considered. As mentioned above, $\mathbf{Y}^{(k)}$ $(k = 1, \ldots, J)$ can denote the data vector of the element-wise trait k from all subjects and the column data vector $\mathbf{Y} = (\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(J)})^{\mathrm{T}}$ in the multivariate GLM model (4.1) is comprised of the element-wise data vectors from all these J elements.

We rearrange the data vector $\mathbf{Y}$ into a subject-by-element data matrix, where there is one row for each subject and one column for each data element (e.g., voxel, surface element, etc) with the notation of

$$\mathbf{Y} = (\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(J)}).$$

Conventional heritability analyses work on a single univariate phenotypic trait, here a single column. Our method then proceeds by computing the correlation coefficient between rows, over phenotypic elements. That is, each twin pair (or subject pair) generates one correlation coefficient r. If we write $\bar{r}_{\mathrm{MZ}}$ and $\bar{r}_{\mathrm{DZ}}$ as the averages of MZ-pair and DZ-pair correlations respectively, then by analogy to Falconer's heritability estimate (3.8), we propose

$$\mathrm{AgHe} = 2 \times (\bar{r}_{\mathrm{MZ}} - \bar{r}_{\mathrm{DZ}}) \tag{4.18}$$

as the aggregate estimate of whole-phenotype heritability for this high-dimensional phenotype and denote this aggregate heritability as "AgHe".

4.3.2 Analytical Derivation of AgHe

As stated previously, the data normalization methods of mean-centering and variance normalization can be applied by calculating the scaled OLS residual for each phenotype using Formula (4.9) before conducting the statistical analysis so as to respectively correct for the possible misestimation due to failure to account for the confounding covariates and phenotypic mean variation, and normalize the data to unit variance to improve the computational stability with a fair comparison between phenotypes. We again assume that all entries of the element-wise phenotypic data vector $\mathbf{Y}^{(k)}$ have zero mean after mean-centering and even have unit variance after further variance-normalization; we neglect any correlation induced by mean centering and variance normalization.

The computation of AgHe is equivalent to computing the pair-wise correlation. We denote the pair-wise correlation for a pair of subjects i and i' with type indices of j and j' to be $r_{ii'}$, which is calculated as the correlation coefficient regarding all element-wise phenotypic data between subjects i and i' with the expression of

$$\begin{aligned} r_{ii'} &= \frac{\sum_{k=1}^{J}(\mathrm{Y}_{i,j}^{(k)} - \bar{\mathrm{Y}}_{i,j})(\mathrm{Y}_{i',j'}^{(k)} - \bar{\mathrm{Y}}_{i',j'})}{\sqrt{\sum_{k=1}^{J}(\mathrm{Y}_{i,j}^{(k)} - \bar{\mathrm{Y}}_{i,j})^2}\sqrt{\sum_{k=1}^{J}(\mathrm{Y}_{i',j'}^{(k)} - \bar{\mathrm{Y}}_{i',j'})^2}} \\ &= \frac{\sum_{k=1}^{J}\sum_{k'=1}^{J}(\mathrm{Y}_{i,j}^{(k)} - \mathrm{Y}_{i,j}^{(k')})(\mathrm{Y}_{i',j'}^{(k)} - \mathrm{Y}_{i',j'}^{(k')})}{\sqrt{\sum_{k=1}^{J}\sum_{k'=1}^{J}(\mathrm{Y}_{i,j}^{(k)} - \mathrm{Y}_{i,j}^{(k')})^2}\sqrt{\sum_{k=1}^{J}\sum_{k'=1}^{J}(\mathrm{Y}_{i',j'}^{(k)} - \mathrm{Y}_{i',j'}^{(k')})^2}}. \end{aligned}$$

Concerning 3 different groups of MZ twins, DZ twins and unrelated individuals (UN), the average pair-wise correlation, denoted as $\bar{r}$ regardless of the group type,

for each group is then approximated and calculated by the formula below:

$$\bar{r} \approx \frac{\mathbb{E}\left[\sum_{k=1}^{J}\sum_{k'=1}^{J}(\mathrm{Y}_{i,j}^{(k)}-\mathrm{Y}_{i,j}^{(k')})(\mathrm{Y}_{i',j'}^{(k)}-\mathrm{Y}_{i',j'}^{(k')})\right]}{\sqrt{\mathbb{E}\left[\sum_{k=1}^{J}\sum_{k'=1}^{J}(\mathrm{Y}_{i,j}^{(k)}-\mathrm{Y}_{i,j}^{(k')})^2\right]}\sqrt{\mathbb{E}\left[\sum_{k=1}^{J}\sum_{k'=1}^{J}(\mathrm{Y}_{i',j'}^{(k)}-\mathrm{Y}_{i',j'}^{(k')})^2\right]}}$$

$$\approx \frac{\mathbb{E}\left[\sum_{k=1}^{J}\sum_{k'=1}^{J}(\mathrm{Y}_{i,j}^{(k)}-\mathrm{Y}_{i,j}^{(k')})(\mathrm{Y}_{i',j'}^{(k)}-\mathrm{Y}_{i',j'}^{(k')})\right]}{\mathbb{E}\left[\sum_{k=1}^{J}\sum_{k'=1}^{J}(\mathrm{Y}_{i,j}^{(k)}-\mathrm{Y}_{i,j}^{(k')})^2\right]},$$

which is exact up to a routine Taylor approximation:

$$\mathbb{E}\left[\frac{\mathrm{B}_1}{\mathrm{B}_2}\right] \approx \frac{\mathbb{E}\left[\mathrm{B}_1\right]}{\mathbb{E}\left[\mathrm{B}_2\right]} \tag{4.19}$$

for random variables B_1 and B_2. Suppose that $f(\mathrm{B}_1,\mathrm{B}_2)=\frac{\mathrm{B}_1}{\mathrm{B}_2}$ and set $\mu_{\mathrm{B}_1}=\mathbb{E}\left[\mathrm{B}_1\right]$ and $\mu_{\mathrm{B}_2}=\mathbb{E}\left[\mathrm{B}_2\right]$. According to the first-order multivariate Taylor series expansion of $f(\cdot)$ about $(\mathrm{B}_1,\mathrm{B}_2)$, a statistical approximation by dropping the remainder is

$$\begin{aligned} f(\mathrm{B}_1,\mathrm{B}_2) &= f(\mu_{\mathrm{B}_1},\mu_{\mathrm{B}_2}) + \sum_{i=1}^{2}(\mathrm{B}_i-\mu_{\mathrm{B}_i})f'_{\mathrm{B}_i}(\mu_{\mathrm{B}_1},\mu_{\mathrm{B}_2}) + \text{Remainder} \\ &\approx f(\mu_{\mathrm{B}_1},\mu_{\mathrm{B}_2}) + \sum_{i=1}^{2}(\mathrm{B}_i-\mu_{\mathrm{B}_i})f'_{\mathrm{B}_i}(\mu_{\mathrm{B}_1},\mu_{\mathrm{B}_2}) \end{aligned}$$

where f'_{B_i} denotes the first-order partial derivative with respect to B_i. Taking the expectation of both sides to get $\mathbb{E}\left[f(\mathrm{B}_1,\mathrm{B}_2)\right] \approx f(\mu_{\mathrm{B}_1},\mu_{\mathrm{B}_2})$, i.e., Equation (4.19).

The utilization of Equations (4.10), (4.11), (4.12) and (4.13) describing the expectations of within-subject cross-phenotype SD's and between-subject cross-phenotype SD's for 3 different groups of subject pairs yields the simplification of the above formula, and then the correlation averages for MZ twins ($\bar{r}_{\mathrm{MZ}}$), DZ twins ($\bar{r}_{\mathrm{DZ}}$) and unrelated individuals ($\bar{r}_{\mathrm{UN}}$) are derived with the expressions of

$$\begin{cases} \bar{r}_{\mathrm{MZ}} = \frac{\overline{wh^2}+\frac{J}{J-1}\overline{wc^2}-\overline{w\{h^2\rho_{\mathrm{G}}\}}}{1-\overline{w\rho_{\mathrm{P}}}}, & \text{for MZ group} \\ \bar{r}_{\mathrm{DZ}} = \frac{\frac{1}{2}\overline{wh^2}+\frac{J}{J-1}\overline{wc^2}-\frac{1}{2}\overline{w\{h^2\rho_{\mathrm{G}}\}}}{1-\overline{w\rho_{\mathrm{P}}}}, & \text{for DZ group} \\ \bar{r}_{\mathrm{UN}} = 0, & \text{for UN group} \end{cases}$$

for mean-centered data, and

$$\begin{cases} \bar{r}_{\mathrm{MZ}} = \frac{\overline{h^2}+\frac{J}{J-1}\overline{c^2}-\overline{\{h^2\rho_{\mathrm{G}}\}}}{1-\overline{\rho_{\mathrm{P}}}}, & \text{for MZ group} \\ \bar{r}_{\mathrm{DZ}} = \frac{\frac{1}{2}\overline{h^2}+\frac{J}{J-1}\overline{c^2}-\frac{1}{2}\overline{\{h^2\rho_{\mathrm{G}}\}}}{1-\overline{\rho_{\mathrm{P}}}}, & \text{for DZ group} \\ \bar{r}_{\mathrm{UN}} = 0, & \text{for UN group} \end{cases}$$

for demeaned and variance-normalized data, where

$$\overline{h^2} = \frac{1}{J}\sum_{k=1}^{J} h_k^2, \quad \overline{c^2} = \frac{1}{J}\sum_{k=1}^{J} c_k^2, \quad \overline{\sigma^2} = \frac{1}{J}\sum_{k=1}^{J} \sigma_k^2,$$

$$\overline{\rho_{\mathrm{P}}} = \frac{2}{J(J-1)}\sum_{k>k'} \rho_{\mathrm{P}}^{(kk')}, \quad \overline{\{h^2\rho_{\mathrm{G}}\}} = \frac{2}{J(J-1)}\sum_{k>k'} h_k h_{k'} \rho_{\mathrm{G}}^{(kk')},$$

$$\overline{wh^2} = \frac{1}{J}\sum_{k=1}^{J} \left(\sigma_r^2 \Big/ \overline{\sigma^2}\right) h_k^2, \quad \overline{wc^2} = \frac{1}{J}\sum_{k=1}^{J} \left(\sigma_r^2 \Big/ \overline{\sigma^2}\right) c_k^2,$$

$$\overline{w\rho_{\mathrm{P}}} = \frac{2}{J(J-1)}\sum_{k>k'} \frac{\sigma_k\sigma_{k'}}{\overline{\sigma^2}} \rho_{\mathrm{P}}^{(kk')},$$

$$\overline{w\{h^2\rho_{\mathrm{G}}\}} = \frac{2}{J(J-1)}\sum_{k>k'} \frac{\sigma_k\sigma_{k'}}{\overline{\sigma^2}} h_k h_{k'} \rho_{\mathrm{G}}^{(kk')}.$$

These expressions show that the unrelated individuals have zero mean pair-wise correlation after data normalization. As expected, the difference between MZ-pair and DZ-pair mean correlations is associated with the genetic factors.

Therefore, the simplified analytical expressions of the AgHe formula (4.18) are derived explicitly as

$$\mathrm{AgHe} = \frac{\overline{wh^2} - \overline{w\{h^2\rho_{\mathrm{G}}\}}}{1-\overline{w\rho_{\mathrm{P}}}}$$

after mean-centering, and

$$\mathrm{AgHe} = \frac{\overline{h^2} - \overline{\{h^2\rho_{\mathrm{G}}\}}}{1-\overline{\rho_{\mathrm{P}}}}$$

after further variance normalization. After data transformation, the MZ-pair and DZ-pair correlation averages differ by half the mean heritability summaries (variance-weighted $\overline{wh^2}$ or unweighted $\overline{h^2}$ heritability averages) over phenotypic elements, though shifted and scaled by $\overline{w\{h^2\rho_{\mathrm{G}}\}}$ and $1-\overline{w\rho_{\mathrm{P}}}$ for mean-centered data, or $\overline{\{h^2\rho_{\mathrm{G}}\}}$ and $1-\overline{\rho_{\mathrm{P}}}$ for demeaned and variance-normalized data. The validity of these analytical results for AgHe is assessed with simulations later by comparing

AgHe measure with the average heritability. As the data transformation method of mean-centering only derives the AgHe formula that takes into account of the element-wise variance variation by using the element-wise variances as the weights, we consider comparing the AgHe measure to the variance-weighted heritability average in addition to the unweighted average.

4.3.3 Hypothesis Testing

If the element-wise multivariate heritability is zero overall, the aggregate heritability must be 0 as well. A non-zero measure of AgHe indicates that there should be some heritable elements. Hence we establish the null hypothesis $\mathrm{H}_0 : \mathrm{AgHe} = 0$ for multivariate hypothesis testing, which is equivalent to assessing the null hypothesis $\mathrm{H}_0 : \bar{r}_{\mathrm{MZ}} = \bar{r}_{\mathrm{DZ}}$ by comparing MZ and DZ groups in terms of their pair-wise correlations. As the MZ and DZ twin pairs are mutually independent, a two-sample t-test can be used to determine whether the means of MZ and DZ pair-wise correlations differ significantly. The formula of this two-sample t-statistic is

$$t = \frac{\bar{r}_{\mathrm{MZ}} - \bar{r}_{\mathrm{DZ}}}{\sqrt{\frac{\mathrm{s}^2_{\mathrm{MZ}}}{(\frac{1}{2}n_{\mathrm{MZ}})} + \frac{\mathrm{s}^2_{\mathrm{DZ}}}{(\frac{1}{2}n_{\mathrm{DZ}})}}},$$

where $\mathrm{s}^2_{\mathrm{MZ}}$ and $\mathrm{s}^2_{\mathrm{DZ}}$ are sample variances on the basis of pair-wise correlations for MZ and DZ groups respectively. The distribution of pair-wise correlation is unknown, and thus the null distribution of this two-sample t-statistic can not be identified. Since the normality assumption can not hold with this test statistic, we consider adopting permutations to construct its empirical distribution and test the null hypothesis, where the permutation approach often has better statistical power than the parametric approaches.

4.3.4 Permutation and Bootstrapping Inferences

As stated in Section 2.5.2, the non-parametric permutation test exists for any test statistic, regardless of whether or not its distribution is known, and can be used for unbalanced designs. The observed data, here the pair-wise correlations, from different twin pairs is labeled as MZ or DZ. Under the null hypothesis for this two-sample testing, these labels are exchangeable, and the rearrangements of the MZ and DZ labels yield permutations. For each permutation, the value of two-sample t-statistic (t) is calculated using the permuted correlations. With the support of permutations, the empirical null distribution of this two-sample t-statistic under the null hypothesis is constructed and the corresponding permutation-based p-value

is calculated in order to test if MZ and DZ groups are significantly different from each other.

In addition to the permutation inference, the bootstrapping inference provides the confidence interval, a supplement to determine rejecting the null hypothesis or not, for mean correlation difference or AgHe for the comparison of MZ and DZ groups. As detailed in Section 2.5.3, the procedure of bootstrapping resampling is to resample twin pairs separately within MZ and DZ groups with replacement, and then the data from the resampled MZ and DZ twin pairs form the bootstrap sample. This resampling process is repeated several times until a sufficiently large number of bootstrap replicates are obtained to yield the bootstrap distribution that can be used for the construction of the confidence intervals. Here we use our proposed joint bootstrap method to construct the bootstrapping confidence intervals for AgHe (see Section 2.5.3).

4.3.5 Simulation-based Evaluations on Data Normalization

We now carry out Monte Carlo simulation to evaluate our analytical measure of AgHe by comparing it to the heritability mean summaries of variance-weighted and unweighted averages using the above-mentioned data normalization methods of mean-centering only and mean-centering & variance-normalization separately.

Simulation Setting

Before conducting the simulations, different simulation settings should be configured firstly. The considered sample is of medium size and assumed to be comprised of $n_{\mathrm{MZ}} = 50$ MZ twins (25 pairs) and $n_{\mathrm{DZ}} = 50$ DZ twins (25 pairs), and $n = 100$ in total. There are entirely 15 settings of the unknown parameters $(A, C, E)^{\mathrm{T}}$, which are shown in Table 3.1. Consider $J = 1000$ element-wise traits with the heritability occurrence of 50% over all these elements. The element-wise phenotypic variance is assumed to vary by the element index from 1 to 1000, the element-wise phenotypic averages are generated from an uniform distribution on the interval $[0, 100]$, and the phenotypic correlation is also set to vary with the element index ranging from 0.091 to 0.985 with zero genetic correlation for each pair of elements. Totally $n_{\mathrm{Rlz}} = 1000$ simulations were performed for each setting.

Results on Data Normalization

Figure 4.6 shows the simulation results, plotting the relative differences (or bias) of AgHe associated with the variance-weighted (top) and unweighted (middle) heritability mean summaries, and the standard deviation for the AgHe estimator (bottom). The standard deviation comparison of AgHe with mean-centering and with further variance normalization in the bottom subfigure shows that these two data normalization methods return fairly similar standard deviation for AgHe with the difference of 0.003–0.01. In the top and middle subfigures, the demeaned and variance-normalized data always obtains smaller magnitudes of the bias than the demeaned data during the comparisons between AgHe and the two heritability mean summaries. While mean-centering & variance-normalization still experiences some bias, these simulation results indicate that the data normalization approach including both mean-centering and variance normalization should be more acceptable than mean-centering only for computing the AgHe estimate to take advantage of the relatively smaller bias between AgHe and heritability averages although the mean-centered data can also provide a good approximation of the heritability mean summaries using AgHe.

4.3.6 Real Data Application

We will demonstrate the ability of this correlation mean difference approach to rank various phenotypes. Using analyses described in the pipelines paper (Glasser et al., 2013), the imaging data of brain structure, function and connectivity were pre-processed in different ways. The proposed aggregate heritability method is used to rank the heritability of totally 22 phenotypes. As a comparison, we use an ACE model fit (Chen et al., 2013) on each of these phenotypes, and then compute the variance-weighted and unweighted mean measures concerning all element-wise heritabilities. We also use permutation and bootstrapping inferences to get p-values and confidence intervals for the aggregate and mean heritability measures.

Data Acquisition

As mentioned above, the HCP project acquired the structural, functional and diffusion MRI data for the inference on brain structure, function and connectivity from a conventional customized Siemens Skyra 3T scanner with high spatial and temporal resolutions and differing distortions, where different pre-processing methods are required to be carried out to achieve optimal results for the later analysis (Glasser et al., 2013). The specially designed minimal pre-processing pipelines, de-

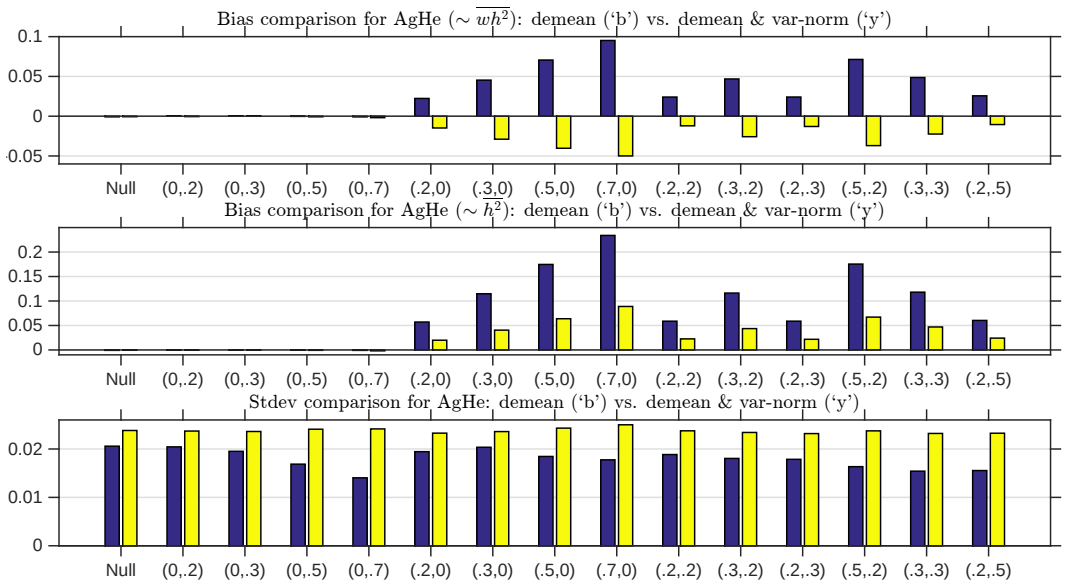


Figure 4.6: The analytical result of applying mean-centering only ('b': blue) is compared with that derived using mean-centering & variance-normalization ('y': yellow). The bias between AgHe and heritability mean summaries of the variance-weighted ($\overline{wh^2}$) and unweighted ($\overline{h^2}$) measures are shown in the top and middle of the figure respectively. The subfigure on the bottom shows the standard deviation of the AgHe measure.

scribed in the HCP pipelines paper (Glasser et al., 2013), were developed to combine and analyze those HCP datasets obtained from multiple MRI neuroimaging modalities together and finally offer the standardized HCP pre-processed data to the community users. These pre-processing methods were applied to the HCP sample including $n = 196$ healthy adults from 105 families, where there exist 32 MZ and 26 DZ twins (i.e., 16 and 13 pairs) and 138 non-twin siblings, and finally resulted in various phenotypic measures (Van Essen et al., 2013). Two stages of inter-subject alignment using the multimodal surface matching (MSM) method (Robinson et al., 2013) were considered to be executed, where stage 1, denoted as "MSM 1", uses the sulcus map to align folding patterns that are consistent across subjects, and stage 2, denoted as "MSM 2", aligns the areal features including the myelin map and resting state networks (RSNs) while ensuring that the average cortical areal size, shape, and position reflect that of the typical subject.

With the use of area ratios and displacements, the areal features were measured and analyzed on the spherical mesh by comparing the local cortical surface area after folding-based registration with that after additional areal-feature-based registration. The surface area ratio is a ratio of surface area of tiles on the sphere using MSM 1 to that using both MSM 1 and MSM 2 (written as "MSM 1+2"), and the displacement is the spatial displacement across the surface of the sphere between MSM 1 and MSM 1+2 alignments. Other structural features were also analyzed including sulcus, myelin, cortical thickness and cortical midthickness coordinates, all after areal-feature-based alignment. Here, sulcus is a folding measure produced by the FreeSurfer pipeline detailed in the HCP pipelines paper (Glasser et al., 2013), myelin maps are as produced in the HCP pipelines paper based on T1-weighted (denoted as "T1w") and T2-weighted (denoted as "T2w") structural scans using the ratio of T1w/T2w, cortical thickness with the surface curvature regressed out is considered to correct for folding so as to reduce the effects of folding on thickness, and cortical midthickness coordinates are the xyz positions of the individual subjects' midthickness vertices after resampling to a standard mesh using MSM alignment (Glasser et al., 2013).

The brain functional connectivity can be derived by simple correlation with mean gray timecourse regressed out (MGTR) or partial correlation (PC). There are two ways of generating individual subject time series, dual regression (DR) and eigen regression (ER). For resting fMRI (rfMRI), the group average resting state PCA series (Smith et al., 2014) was decomposed into 100 ICA components and regressed

into individual subject resting state time series. After removing the mean gray matter signal, time courses for each spatial component were correlated to generate parcellated functional connectomes (100×100). In the task analyses, the ICA components based on ICA decomposition were regressed into the task spatial maps, the individual subjects' task fMRI (tfMRI) contrast activation z-statistic maps, to produce parcellated task maps with two dimensionalities of 100 and 200, where all 7 tasks and totally 86 HCP task contrasts were used.

Results

Tables 4.3, 4.4 and 4.5 report the estimates, 95% bootstrapping confidence intervals with 1000 bootstrap replicates and permutation-based p-values derived using 1000 permutations of heritability mean summaries and AgHe for the resulting 6 HCP structural phenotypic measures, 12 HCP phenotypes related to functional connectivity and measured with rfMRI, and 4 HCP task-related phenotypes measured using tfMRI for all tasks respectively. Comparing the two heritability summaries we find the variance-weighted average $\overline{wh^2}$ is generally higher than the unweighted mean $\overline{h^2}$, suggesting that the elements with more variability over subjects are more heritable. Concerning these 22 HCP phenotypes including brain structure, function and connectivity, we find evidence for heritability using those three heritability summary measures in all phenotypes except a high-dimensional measure of functional connectivity that was only significant for AgHe, and a structural measure of myelin that was only significant using the unweighted mean. The AgHe measure closely follows the results of unweighted and variance-weighted averages, and thus is comparable to those unweighted and variance-weighted heritability mean summaries, supporting its utility as a quick screening heritability estimate.

In the upper part of Figure 4.7, the aggregate measure of heritability is compared to both variance-weighted (right) and unweighted (left) means of traditional heritability estimates in ranking the above-mentioned 22 phenotypes. There is a monotonic relationship found between AgHe and these heritability mean summaries in terms of the estimates. We also find the relationship between AgHe and variance-weighted mean is stronger than that between AgHe and unweighted mean. In the lower part of Figure 4.7, both heritability mean summaries are found to obtain strong significance for most phenotypes, so does the AgHe measure. Notably, there is good concordance in terms of the p-values between mean summaries and AgHe.

The comparison between mean summaries and AgHe in Tables 4.3, 4.4 and 4.5

and Figure 4.7 implies that this extremely fast aggregate measure of heritability is highly similar to the traditional but more computationally intensive mean heritability summaries obtained by fitting ACE model. This simple correlation mean difference approach provides a simple omnibus test for multivariate heritability inference with a simple summary measure of AgHe that can be treated as a fast estimate of the whole-brain heritability.

4.4 Summary of the Chapter

In this chapter, we have generalized the univariate LR-SD method to the bivariate case for estimating genetic and phenotypic correlations, and we illustrate this bivariate method with a real data application. Although the estimation accuracy for bivariate LR-SD is not exactly as anticipated in simulations, the negligible MSE and controlled FPR for the ERV estimator imply it as an adequate test statistic. Moreover, we also have investigated a simple and rapid correlation mean difference approach for multivariate heritability analysis to assess the genetic differences between MZ and DZ twins. There is a close relationship found between the resulting aggregate heritability and the heritability mean summaries, suggesting AgHe to be a fast screening measure of the whole-phenotype heritability for a high-dimensional phenotype.

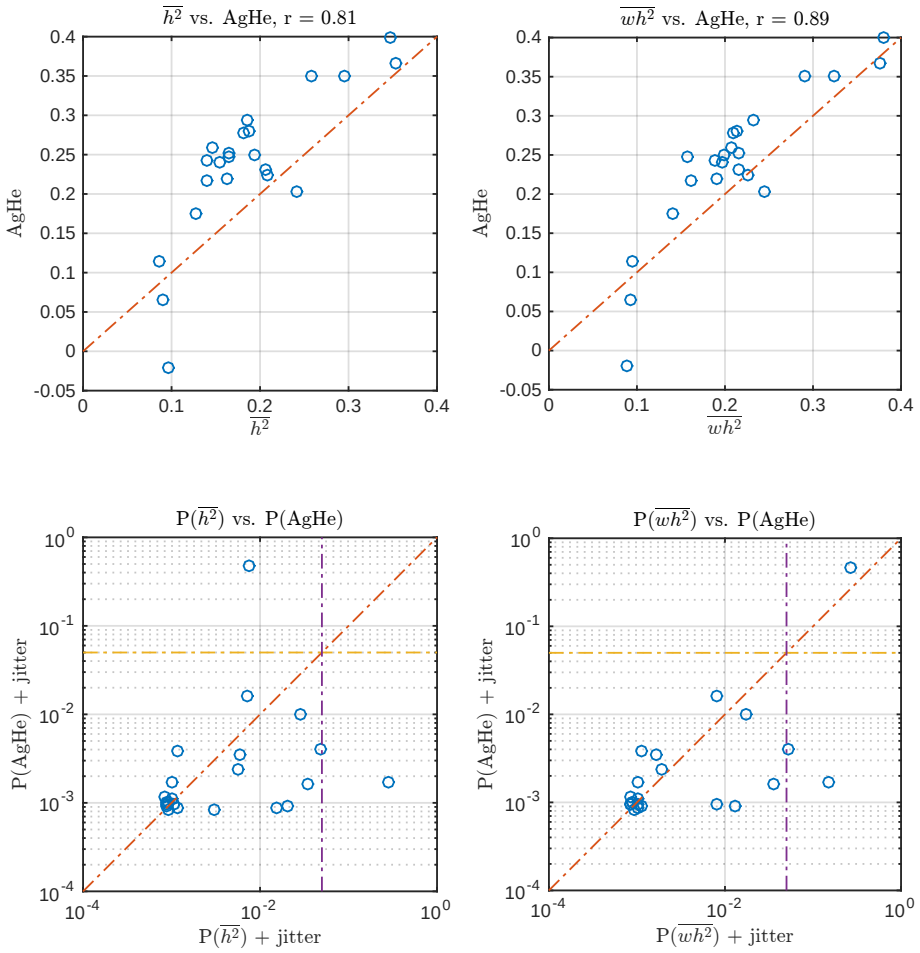


Figure 4.7: The estimate (upper) and p-value (lower) comparisons between AgHe and the variance-weighted ($\overline{wh^2}$) and unweighted ($\overline{h^2}$) mean summaries. The uniformly distributed jitter with the distribution of $\mathbb{U}(0.8, 1.2)$ was used to allow the visualization of the many p-values near 10^{-3}.

Table 4.3: The estimates, permutation-based p-values derived using 1000 permutations and 95% CIs with 1000 bootstrap replicates for the unweighted ($\overline{h^2}$) and variance-weighted ($\overline{wh^2}$) heritability mean summaries and AgHe for the 6 HCP structural phenotypic measures.

	Estimate	**95% CI**	**P-value**
	Surface Area Ratio		
$\overline{h^2}$	0.181	(0.161, 0.200)	0.001
$\overline{wh^2}$	0.210	(0.183, 0.233)	0.001
AgHe	0.278	(0.179, 0.366)	0.001
	Spatial Displacement		
$\overline{h^2}$	0.347	(0.291, 0.374)	0.001
$\overline{wh^2}$	0.381	(0.317, 0.411)	0.001
AgHe	0.399	(0.236, 0.562)	0.001
	Sulcus (FreeSurfer)		
$\overline{h^2}$	0.164	(0.145, 0.188)	0.001
$\overline{wh^2}$	0.157	(0.140, 0.177)	0.001
AgHe	0.248	(0.185, 0.307)	0.001
	Myelin (T1w/T2w)		
$\overline{h^2}$	0.096	(0.078, 0.121)	0.009
$\overline{wh^2}$	0.089	(0.064, 0.127)	0.309
AgHe	-0.020	(-0.223, 0.191)	0.573
	Folding-corrected Cortical Thickness		
$\overline{h^2}$	0.091	(0.076, 0.117)	0.042
$\overline{wh^2}$	0.093	(0.078, 0.119)	0.043
AgHe	0.066	(0.027, 0.105)	0.002
	3D Cortical Midthickness Coordinates		
$\overline{h^2}$	0.194	(0.171, 0.221)	0.001
$\overline{wh^2}$	0.199	(0.176, 0.223)	0.001
AgHe	0.249	(0.136, 0.359)	0.001

Table 4.4: The estimates, permutation-based p-values derived using 1000 permutations and 95% CIs with 1000 bootstrap replicates for the unweighted ($\overline{h^2}$) and variance-weighted ($\overline{wh^2}$) heritability mean summaries and AgHe for the 12 HCP phenotypes related to functional connectivity. (to be continued on the next page)

	Estimate	**95% CI**	**P-value**
	rfMRI (MGTR, DR, dim: 100, MSM 1+2)		
$\overline{h^2}$	0.353	(0.273, 0.397)	0.001
$\overline{wh^2}$	0.375	(0.289, 0.423)	0.001
AgHe	0.367	(0.148, 0.587)	0.004
	rfMRI (MGTR, DR, dim: 200, MSM 1+2)		
$\overline{h^2}$	0.296	(0.234, 0.335)	0.001
$\overline{wh^2}$	0.325	(0.255, 0.369)	0.001
AgHe	0.350	(0.167, 0.533)	0.002
	rfMRI (PC (ridgep = 1), DR, dim: 100, MSM 1+2)		
$\overline{h^2}$	0.259	(0.211, 0.286)	0.001
$\overline{wh^2}$	0.290	(0.237, 0.318)	0.001
AgHe	0.350	(0.219, 0.483)	0.001
	rfMRI (PC (ridgep = 1), DR, dim: 200, MSM 1+2)		
$\overline{h^2}$	0.186	(0.161, 0.212)	0.003
$\overline{wh^2}$	0.232	(0.198, 0.257)	0.001
AgHe	0.295	(0.184, 0.394)	0.001
	rfMRI (PC (icov), DR, dim: 100, MSM 1+2)		
$\overline{h^2}$	0.128	(0.114, 0.142)	0.001
$\overline{wh^2}$	0.141	(0.128, 0.154)	0.001
AgHe	0.175	(0.124, 0.227)	0.001
	rfMRI (PC (icov), DR, dim: 200, MSM 1+2)		
$\overline{h^2}$	0.086	(0.075, 0.107)	0.001
$\overline{wh^2}$	0.095	(0.084, 0.114)	0.001
AgHe	0.115	(0.078, 0.151)	0.001

	Estimate	95% CI	P-value
	rfMRI (PC (ridgep = 1), ER, dim: 100, MSM 1+2)		
$\overline{h^2}$	0.206	(0.160, 0.245)	0.042
$\overline{wh^2}$	0.215	(0.164, 0.256)	0.045
AgHe	0.231	(0.075, 0.387)	0.005
	rfMRI (PC (ridgep = 1), ER, dim: 200, MSM 1+2)		
$\overline{h^2}$	0.162	(0.132, 0.196)	0.239
$\overline{wh^2}$	0.191	(0.154, 0.228)	0.126
AgHe	0.219	(0.101, 0.340)	0.002
	rfMRI (MGTR, ER, dim: 100, MSM 1+2)		
$\overline{h^2}$	0.241	(0.187, 0.279)	0.008
$\overline{wh^2}$	0.244	(0.187, 0.287)	0.009
AgHe	0.203	(0.004, 0.398)	0.016
	rfMRI (MGTR, ER, dim: 200, MSM 1+2)		
$\overline{h^2}$	0.208	(0.156, 0.258)	0.029
$\overline{wh^2}$	0.226	(0.167, 0.281)	0.018
AgHe	0.225	(0.049, 0.402)	0.009
	rfMRI (PC (ridgep = 1), DR, dim: 100, MSM 1)		
$\overline{h^2}$	0.147	(0.128, 0.167)	0.007
$\overline{wh^2}$	0.207	(0.174, 0.231)	0.002
AgHe	0.260	(0.119, 0.421)	0.003
	rfMRI (PC (ridgep = 1), DR, dim: 200, MSM 1)		
$\overline{h^2}$	0.188	(0.163, 0.211)	0.001
$\overline{wh^2}$	0.213	(0.184, 0.233)	0.001
AgHe	0.280	(0.187, 0.370)	0.001

Table 4.5: The estimates, permutation-based p-values derived using 1000 permutations and 95% CIs with 1000 bootstrap replicates for the unweighted ($\overline{h^2}$) and variance-weighted ($\overline{wh^2}$) heritability mean summaries and AgHe for the 4 task-related HCP phenotypes.

	Estimate	**95% CI**	**P-value**
	Parcellated tfMRI for all tasks (dim: 100, MSM 1+2)		
$\overline{h^2}$	0.154	(0.132, 0.181)	0.018
$\overline{wh^2}$	0.196	(0.160, 0.234)	0.015
AgHe	0.240	(0.076, 0.429)	0.001
	Parcellated tfMRI for all tasks (dim: 200, MSM 1+2)		
$\overline{h^2}$	0.139	(0.120, 0.161)	0.023
$\overline{wh^2}$	0.188	(0.156, 0.229)	0.009
AgHe	0.243	(0.104, 0.396)	0.001
	Parcellated tfMRI for all tasks (dim: 100, MSM 1)		
$\overline{h^2}$	0.164	(0.142, 0.188)	0.006
$\overline{wh^2}$	0.216	(0.177, 0.241)	0.002
AgHe	0.252	(0.092, 0.420)	0.002
	Parcellated tfMRI for all tasks (dim: 200, MSM 1)		
$\overline{h^2}$	0.147	(0.128, 0.167)	0.007
$\overline{wh^2}$	0.207	(0.174, 0.231)	0.002
AgHe	0.260	(0.119, 0.421)	0.003

Chapter 5

Conclusion and Future Work

This chapter begins with summarising our work described in Chapter 3 and Chapter 4. We then focus on the discussion of the potential future research direction.

5.1 Conclusion and Discussion

In this thesis, we developed a series of approaches for the inference on the additive genetic factors including heritability and genetic correlation, which have been and are currently being used in multiple projects.

In Chapter 3, we proposed a Frequentist ReML by modifying the conventional ReML estimation approach with the use of logarithm parameterization, reparameterization, Fisher scoring algorithm and line search, and two linear regression methods new to the neuroimaging field by constructing the mean structure of the linear regression models with squared differences of paired observations and estimating the heritability with LR-SD and LR-SD ReML. A permutation-based heritability inference approach by embedding the LR-SD method in a permutation framework was also developed for both voxel- and cluster-wise inferences.

This permutation inference allows us to perform more exact heritability inference at each voxel to control the family-wise error rate, and also to consider the alternate cluster-wise imaging statistics, such as cluster size and cluster mass, and the whole-image summary statistics. Our use of this fast, accurate and non-iterative LR-SD method (free of any convergence issues) makes these spatially informed statistics more accessible. For equivalent family-wise error rates, the cluster-wise approach was found to have higher sensitivity, and thus more powerful than the voxel-wise

method in power evaluations, which demonstrates the importance of such spatial statistics and the need for permutation inference to take advantage of these cluster statistics. With only few weak assumptions, permutation inference is a feasible alternative to the parametric approaches, and is even preferable in small sample problems or when the stronger assumptions of the parametric approaches cannot be met (Nichols and Holmes, 2001).

In Chapter 4, our newly proposed univariate LR-SD method is generalized to the bivariate case, where the bivariate linear regression models are built by the expectations of the difference between cross-phenotype squared differences, and combining the univariate and bivariate LR-SD methods form the estimation of the correlation parameters. A fast multivariate heritability inference is also established with the use of pair-wise correlation mean difference approach and two-sample t-test, where the aggregate measure of heritability is investigated as a fast screening estimate of the whole-phenotype heritability for a high-dimensional phenotype.

Although the estimation accuracy for the bivariate LR-SD method is not that satisfying as expected, the simulation results for the ERV estimator reveals its property of insignificant MSE magnitude with the false positive error rate under control, suggesting the suitability of this chosen test statistic by reducing the computational complexity. The use of this fast ERV test statistic allows the speedy implementation of permutation test. Concerning all the element-wise traits, the correlation mean difference approach compares MZ and DZ twins for the investigation of the additive genetic influences that lead to the difference between MZ and DZ mean correlations. A non-zero aggregate estimate of heritability can be an excellent indication of the existence of heritable elements, and a zero aggregate measure implies that there are no heritable elements over the brain.

5.2 Future Perspectives

Although the permutation-based heritability inference has increased the statistical power with the use of cluster-wise approach, the power of this non-parametric test is still low in an absolute sense, less than 80%. The feature that the adjacent voxels in the brain are tend to be structurally and functionally more homologous can be utilized by considering the spatially modeling approaches to gather the information from the neighboring regions and improve the sensitivity of the test.

Neuroimaging data is typified by spatial smoothness, both in the signal and noise. Correlation between neighboring voxels' random errors is due to a number of effects: imperfect imaging devices, where the point-spread-function extends beyond the voxel boundaries; spatial transformations, like head motion correction or inter-subject warping, where interpolation induces spatial correlation; and unmodeled individual differences that possess spatial structure will also induce spatially correlated errors (Poldrack et al., 2011). Spatially extended signal is expected due to the modular structure of the brain (functional specialization) (Penny et al., 2007). As most users find improved empirical results with some amount of spatial smoothing in pre-processing to average out the white noise, spatial smoothing has become a ubiquitous feature of neuroimaging data analysis. The analogical spatial regularization through smoothing of the voxel-wise statistic images can also assist in the statistical analysis to improve the validity and sensitivity of our tests, which will be discussed in the future.

Our presented work in the thesis was based on twin design using the traditional ACE model, however, the statistical power for different experimental designs were examined and the family design including twins and one or two additional non-twin siblings, as opposed to the classical twin design, substantially increases the statistical power (Posthuma and Boomsma, 2000). Therefore, our next task is the extension and generalization of the above-mentioned approaches for the extended family design with various compositions of relatives having different degrees of relatedness that need to be adapted to.

Bibliography

C.F. Beckmann, M. Jenkinson, and S.M. Smith. General multilevel linear modeling for group analysis in fmri. *NeuroImage*, 20(2):1052–1063, 2003.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001.

G.A.M. Blokland, K.L. McMahon, P.M. Thompson, N.G. Martin, G.I. de Zubicaray, and M.J. Wright. Heritability of working memory brain activation. *The Journal of Neuroscience*, 31(30):10882–10890, 2011.

T.D. Cannon, J. Kaprio, J. Lönnqvist, M. Huttunen, and M. Koskenvuo. The genetic epidemiology of schizophrenia in a finnish twin cohort. *Archives of General Psychiatry*, 55(1):67–74, 1998.

D.P. Chakraborty and L.H. Winter. Free-response methodology: Alternate analysis and a new observer-performance experiment. *Radiology*, 174:873–881, 1990.

X. Chen, G.A. Blokland, L. Strike, and T.E. Nichols. Voxel-wise and cluster-based heritability inferences of fmri data, 2013. Poster presented at 19th Annual Meeting of the Organization for Human Brain Mapping (OHBM), June 16–20, Seattle, USA.

E.H. Chudle. *Inside Your Brain*. Chelsea House Publishers, 2007.

J.A. Coan and J.J.B. Allen. *Handbook of Emotion Elicitation and Assessment*. Oxford University Press, 2007.

T.J. DiCiccio and B. Efron. Bootstrap confidence intervals. *Statistical Science*, 11 (3):189–228, 1996.

A. Dominicus, A. Skrondal, H.K. Gjessing, N.L. Pedersen, and J. Palmgren. Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics*, 36(2):331–340, 2006.

D.S. Falconer and T.F.C. Mackay. *Introduction to Quantitative Genetics (4th Edition)*. Longman, 1996.

N. Filippini, A. Rao, S. Wetten, R.A. Gibson, M. Borrie, D. Guzman, A. Kertesz, I. Loy-English, J. Williams, T.E. Nichols, B. Whitcher, and P.M. Matthews. Anatomically-distinct genetic associations of apoe $\epsilon 4$ allele load with regional cortical atrophy in alzheimer's disease. *NeuroImage*, 44(3):724–728, 2008.

R.S.J. Frackowiak, K.J. Friston, C.D. Frith, R.J. Dolan, C.J. Price, S. Zeki, J.T. Ashburner, and W.D. Penny. *Human Brain Function (2nd Edition)*. Academic Press, 2004.

K.J. Friston, D.E. Glaser, R.N. Henson, S. Kiebel, C. Phillips, and J. Ashburner. Classical and bayesian inference in neuroimaging: Applications. *NeuroImage*, 16 (2):484–512, 2002a.

K.J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner. Classical and bayesian inference in neuroimaging: Theory. *NeuroImage*, 16(2):465–483, 2002b.

D.C. Glahn, P.M. Thompson, and J. Blangero. Neuroimaging endophenotypes: Strategies for finding genes influencing brain structure and function. *Human Brain Mapping*, 28(6):488–501, 2007.

D.C. Glahn, J.E. Curran, A.M. Winkler, M.A. Carless, J.W. Kent Jr., J.C. Charlesworth, M.P. Johnson, H.H.H. Göring, S.A. Cole, T.D. Dyer, E.K. Moses, R.L. Olvera, P. Kochunov, R. Duggirala, P.T. Fox, L. Almasy, and J. Blangero. High dimensional endophenotype ranking in the search for major depression risk genes. *Biological Psychiatry*, 71(1):6–14, 2012.

M.F. Glasser, S.N. Sotiropoulos, J.A. Wilson, T.S. Coalson, B. Fischl, J.L. Andersson, J. Xu, S. Jbabdi, M. Webster, J.R. Polimeni, D.C. Van Essen, M. Jenkinson M, and WU-Minn HCP Consortium. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

J.R. Gleason. Algorithms for balanced bootstrap simulations. *The American Statistician*, 42(4):263–266, 1988.

B.L. Goldstein and D.N. Klein. A review of selected candidate endophenotypes for depression. *Clinical Psychology Review*, 34(5):417–427, 2014.

I.H. Gotlib and J.P. Hamilton. Bringing genetics back to psychiatric endophenotypes. *Biological Psychiatry*, 71(1):2–3, 2012.

I.I. Gottesman and T.D. Gould. The endophenotype concept in psychiatry: Etymology and strategic intentions. *The American Journal of Psychiatry*, 160(4): 636–645, 2003.

L.W. Grimes and W.R. Harvey. Estimation of genetic variances and covariances using symmetric differences squared. *Journal of Animal Science*, 50(4):634–644, 1980.

A.R. Hariri, E.M. Drabant, and D.R. Weinberger. Imaging genetics: Perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. *Biological Psychiatry*, 59(10):888–897, 2006.

D.A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358): 320–338, 1977.

R.H. Hashemi, W.G. Bradley, and C.J. Lisanti. *MRI: The Basics (3rd Edition)*. Lippincott Williams & Wilkins, 2010.

N. Jahanshad, P.V. Kochunov, E. Sprooten, R.C. Mandl, T.E. Nichols, L. Almasy, J. Blangero, R.M. Brouwer, J.E. Curran, G.I. de Zubicaray, R. Duggirala, P.T. Fox, L.E. Hong, B.A. Landman, N.G. Martin, K.L. McMahon, S.E. Medland, B.D. Mitchell, R.L. Olvera, C.P. Peterson, J.M. Starr, J.E. Sussmann, A.W. Toga, J.M. Wardlaw, M.J. Wright, H.E. Hulshoff Pol, M.E. Bastin, A.M. McIntosh, I.J. Deary, P.M. Thompson, and D.C. Glahn. Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the enigma-dti working group. *NeuroImage*, 81:455–469, 2013.

R.I. Jennrich and P.F. Sampson. Newton-raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, 18(1):11–17, 1976.

W. Karush. Minima of functions of several variables with inequalities as side constraints. Master's thesis, Department of Mathematics, University of Chicago, Chicago, IL, USA, 1939.

H.W. Kuhn and A.W. Tucker. Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*, pages 481–492. Berkeley: University of California Press, 1951.

C.L. Lawson and R.J. Hanson. *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics, 1987.

A.D. Lee, N. Leporé, J. de Leeuw, C.C. Brun, M. Barysheva, K.L. McMahon, G.I. de Zubicaray, N.G. Martin, M.J. Wright, and P.M. Thompson. Multivariate variance-components analysis in dti. *IEEE International Symposium on Biomedical Imaging*, pages 1157–1160, 2010.

M.A. Lindquist, J. Spicer, I. Asllani, and T.D. Wager. Estimating and testing variance components in a multi-level glm. *Neuroimage*, 59(1):490–501, 2012.

N.K. Logothetis. The underpinnings of the bold functional magnetic resonance imaging signal. *The Journal of Neuroscience*, 23(10):3963–3971, 2003.

Y. Luo and R. Duraiswami. Efficient parallel non-negative least squares on multicore architectures. *SIAM J. on Scientific Computing*, 33(5):2848–2863, 2011.

P. Mansfield. Multi-planar image formation using nmr spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):150–158, 1977.

P. McGuffin, F. Rijsdijk, M. Andrew, P. Sham, R. Katz, and A. Cardno. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Archives of General Psychiatry*, 60(5):497–502, 2003.

P. McGuffin, M.J. Owen, and I.I. Gottesman. *Psychiatric Genetics and Genomics*. Oxford University Press, 2004.

K. Mishchenko, S. Holmgren, and L. Roïnegård. Newton-type methods for reml estimation in genetic analysis of quantitative traits. *Journal of Computational Methods in Sciences and Engineering*, 8(1):53–67, 2008.

S. Mori, K. Oishi, H. Jiang, L. Jiang, X. Li, K. Akhter, K. Hua, A.V. Faria, A. Mahmood, R. Woods, A.W. Toga, G.B. Pike, P.R. Neto, A. Evans, J. Zhang, H. Huang, M.I. Miller, P. van Zijl, and J. Mazziotta. Stereotaxic white matter atlas based on diffusion tensor imaging in an icbm template. *Neuroimage*, 40(2): 570–582, 2008.

J.A. Muford and T.E. Nichols. Modeling and inference of multisubject fmri data. *Engineering in Medicine and Biology Magazine, IEEE*, 25(2):42–51, 2006.

T.E. Nichols and S. Hayasaka. Controlling the familiwise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, 12 (5):419–446, 2003.

T.E. Nichols and A.P. Holmes. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1):1–25, 2001.

T.E. Nichols, K. Friston, J. Roiser, and E. Viding. Improving heritability estimates with restricted maximum likelihood (reml), 2009. Poster presented at 15th Annual Meeting of the Organization for Human Brain Mapping (OHBM), June 18–23, San Francisco, CA, USA.

J. Nocedal and S.J. Wright. *Numerical Optimization (2nd Edition)*. Springer, 1999.

S. Ogawa, T.M. Lee, A.S. Nayak, and P. Glynn. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, 14(1):68–78, 1990.

M.R. Osborne. Fisher's method of scoring. *International Statistical Review/Revue Internationale de Statistique*, 60(1):99–117, 1992.

H.D. Patterson and R. Thompson. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3):545–554, 1971.

W.D. Penny, K.J. Friston, J.T. Ashburner, S.J. Kiebel, and T.E. Nichols. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2007.

R.A. Poldrack, J.A. Mumford, and T.E. Nichols. *Handbook of Functional MRI Analysis*. Cambridge University Press, 2011.

D. Posthuma and D.I. Boomsma. A note on the statistical power in extended twin designs. *Behavior Genetics*, 30(2):147–158, 2000.

M.E. Raichle. The brain's dark energy. *Science*, 314(5803):1249–1250, 2006.

F.V. Rijsdijk and P.C. Sham. Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics*, 3(2):119–133, 2002.

E.C. Robinson, S. Jbabdi, J.L.R. Andersson, S.M. Smith, M.F. Glasser, D.C. Van Essen, G.C. Burgess, M.P. Harms, D.M. Barch, and M. Jenkinson. Multimodal surface matching: Fast and generalisable cortical registration using discrete optimisation. In *Information Processing in Medical Imaging (IPMI 2013)*, volume 7917 of *Lecture Notes in Computer Science*, pages 475–486, 2013.

S.R. Searle. Phenotypic, genetic and environmental correlations. *Biometrics*, 17(3): 474–480, 1961.

S.G. Self and K. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.

S.M. Smith and T.E. Nichols. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1):83–98, 2009.

S.M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T.E. Nichols, C.E. Mackay, K.E. Watkins, O. Ciccarelli, M.Z. Cader, P.M. Matthews, and T.E. Behrens. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31(4):1487–1505, 2006.

S.M. Smith, A. Hyvärinen, G. Varoquauxc, K.L. Millera, and C.F. Beckmann. Group-pca for very large fmri datasets. *NeuroImage*, 101:738–749, 2014.

J.L. Stein, X. Hua, S. Lee, A.J. Ho, A.D. Leow, A.W. Toga, A.J. Saykin, L. Shen, T. Foroud, N. Pankratz, M.J. Huentelman, D.W. Craig, J.D. Gerber, A.N. Allen, J.J. Corneveaux, B.M. DeChairo, S.G. Potkin, M.W. Weiner, P.M. Thompson, and the Alzheimer's Disease Neuroimaging Initiative. Voxelwise genome-wide association study (vgwas). *NeuroImage*, 53(3):1160–1174, 2010.

K. Tairyan and J. Illes. Imaging genetics and the power of combined technologies: A perspective from neuroethics. *Neuroscience*, 164(1):7–15, 2009.

D.C. Van Essen, S.M. Smith, D.M. Barch, T.E. Behrens, E. Yacoub, K. Ugurbil K, and WU-Minn HCP Consortium. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

D. Weishaupt, V.D. Köchli, and B. Marincek. *How does MRI work? An Introduction to the Physics and Function of Magnetic Resonance Imaging (2nd Edition)*. Springer, 2006.

S.S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

A.M. Winkler, G.R. Ridgway ad M.A. Webster, S.M. Smith, and T.E. Nichols. Permutation inference for the general linear model. *Neuroimage*, 92(100):381–C397, 2014.

P. Wolfe. Convergence conditions for ascent methods. *SIAM Review*, 11(2):226–235, 1969.

P. Wolfe. Convergence conditions for ascent methods. ii: Some corrections. *SIAM Review*, 13(2):185–188, 1971.

I. Wright, P. Sham, R. Murray, D.R. Weinberger, and E.T. Bullmore. Genetic contributions to regional variability in human brain structure: Methods and preliminary results. *Neuroimage*, 17(1):256–271, 2002.

D. Yekutieli and Y. Benjamini. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, 82(1–2):171–196, 1999.

D. Zhang and X. Lin. Variance component testing in generalized linear mixed models for longitudinal/clustered data and other related topics. *Random Effect and Latent Variable Model Selection*, 192:19–36, 2008.